Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematics

Institute of IIT Kanpur

Lecture: 35

Hello everyone, this is the fifth lecture of week 7. Recall that in the previous four lectures we learned about the quadratic penalty method. We learned how to construct the quadratic penalty function for problems with only equality constraints and for problems with only inequality constraints. We also provided sample code for examples with only equality constraints and only inequality constraints.

To complete this topic, we will now address what to do when you have both equality and inequality constraints. This was the original problem we considered initially. If you recall, we have the problem of minimizing f(x) subject to p inequality constraints and m equality constraints.

Notation: $\mathbf{x_{k+1}= x^{k+1}}$, $\boldsymbol{\gamma_{k+1}= \gamma^{k+1}}$, $\boldsymbol{\nabla^2=\nabla^2}$

When you have this general problem, the quadratic penalty function for both equality and inequality constraints is defined as:

$Q(x, \gamma) = f(x) + \gamma \, [\sum_{j=1}^{m} (h_j(x))^2 + \sum_{i=1}^{p} (g_i(x)_+)^2]$

Recall that $g_i(x)_+$ means $\max(0, g_i(x))$. This is nothing new—you simply need to add up the penalties from all constraints. Whenever a particular equality constraint is not satisfied, you impose a penalty of γ for that constraint. Similarly, whenever an inequality constraint is violated (i.e., $g_i(x) > 0$), you impose a penalty of γ based on the value of $g_i(x)_+$.

When writing code for problems with both equality and inequality constraints, the approach is straightforward. For problems with only inequality constraints, you defined f, g, and ∇g.

For problems with only equality constraints, you defined f, h, and ∇h. For problems with both, you need to define h, ∇h, g, and ∇g, and then define Q, ∇Q, and the Hessian H accordingly.

In the expression for Q, you will have $h(x) \cdot h(x) + (g(x) > 0) * g(x) \cdot g(x)$.

Similarly, in $\nabla Q(x)$, you will have $\nabla f(x)$ plus the contributions from the equality and inequality constraints. In the Hessian H, you will have the Hessian of f, plus $2\gamma \, \nabla h(x) \, \nabla h(x)^T$, and also $2\gamma \, h(x) \, \nabla^2 h(x)$.

In all the examples we considered, the Hessian of h was zero, which is why that term did not appear. However, for a general problem, you must write Q, ∇Q, and H carefully.

I could work out an example, but given that we have already covered examples for equality and inequality constraints separately, it is not necessary. Instead, we will proceed to the analysis of the quadratic penalty method.

Recall the algorithm:

1. Initialize x_0 and set $k = 0$.

2. Set the tolerance and choose a sequence $\gamma_0, \gamma_1, \gamma_2, \ldots$ that is increasing and diverges to infinity.

3. While $|Q(x_k, \gamma_k) - f(x_k)| >$ tolerance:

   - Find x that minimizes $Q(x, \gamma_k)$

   - Update $x_{k+1} = x$

   - Increment k

4. Output x_k as the minimizer.

A natural question is why this method converges to the correct answer. We have seen examples where it converges, but is there a proof? Can we understand the properties of the method and what happens in each iteration?

I will present a **Theorem** that describes the behaviour of the method. For the quadratic penalty method, we have three results:

1. $Q(x_k, \gamma_k) \leq Q(x_{k+1}, \gamma_{k+1})$

2. The penalty $p(x_k) \geq p(x_{k+1})$

3. $f(x_k) \leq f(x_{k+1})$

This theorem compares successive iterations of the outer loop. The value of Q increases, the penalty decreases, and the objective function f increases as we proceed through the iterations.

Recall that $Q(x, \gamma) = f(x) + \gamma\, p(x)$, where $p(x) = \sum_j (h_j(x))^2 + \sum_i (g_i(x)_+)^2$.

The theorem states that Q increases, f increases, but p decreases in successive iterations.

Proof: First, note that x_k = argmin over x of $Q(\bar{x}, \gamma_k)$.

Therefore, $Q(x_k, \gamma_k) \leq Q(x, \gamma_k)$ for any other x.

For the first result:

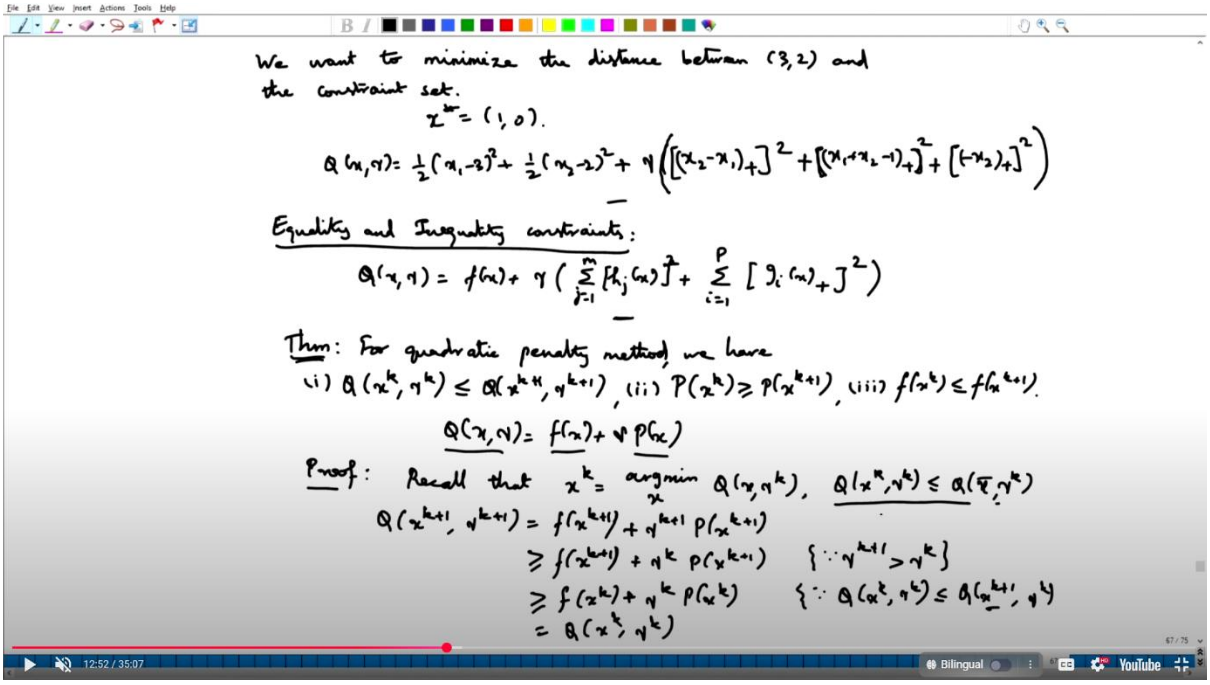
$Q(x_{k+1}, \gamma_{k+1}) = f(x_{k+1}) + \gamma_{k+1}\, p(x_{k+1})$

$\geq f(x_{k+1}) + \gamma_k\, p(x_{k+1})$ since ($\gamma_{k+1} > \gamma_k$)

$\geq f(x_k) + \gamma_k\, p(x_k)$ Since $Q(x_k, \gamma_k) \leq Q(x_{k+1}, \gamma_{k+1})$

$\geq Q(x_k, \gamma_k)$

(Refer Slide Time 12:52)



The first inequality holds because $\gamma_{k+1} > \gamma_k$ and $p(x_{k+1}) \ge 0$.

The second inequality holds because x_k minimizes $Q(x, \gamma_k)$.

For the second result, $p(x_k) \ge p(x_{k+1})$:

From the minimization property:

$f(x_k) + \gamma_k\, p(x_k) \le f(x_{k+1}) + \gamma_k\, p(x_{k+1})$

$f(x_{k+1}) + \gamma_{k+1}\, p(x_{k+1}) \le f(x_k) + \gamma_{k+1}\, p(x_k)$

Adding these two inequalities:

$f(x_k) + f(x_{k+1}) + \gamma_k\, p(x_k) + \gamma_{k+1}\, p(x_{k+1}) \le f(x_k) + f(x_{k+1}) + \gamma_k\, p(x_{k+1}) + \gamma_{k+1}\, p(x_k)$

Canceling $f(x_k) + f(x_{k+1})$ from both sides:

$\gamma_k\, p(x_k) + \gamma_{k+1}\, p(x_{k+1}) \le \gamma_k\, p(x_{k+1}) + \gamma_{k+1}\, p(x_k)$

Rearranging:

$(\gamma_k - \gamma_{k+1})(p(x_k) - p(x_{k+1})) \le 0$

Since $\gamma_{k+1} > \gamma_k$, we have $\gamma_k - \gamma_{k+1} < 0$, so $p(x_k) - p(x_{k+1}) \ge 0$, hence $p(x_k) \ge p(x_{k+1})$.

For the third result, $f(x_k) \le f(x_{k+1})$:

From $f(x_k) + \gamma_k\, p(x_k) \le f(x_{k+1}) + \gamma_k\, p(x_{k+1})$

Rearranging:

$f(x_k) - f(x_{k+1}) \leq \gamma_k (p(x_{k+1}) - p(x_k))$

Since $p(x_k) \geq p(x_{k+1})$, the right-hand side is ≤ 0, so $f(x_k) \leq f(x_{k+1})$.

This completes the proof.

To explain these results: the objective function f increases in successive iterations. This might be surprising because we want to minimize f(x). However, all the points x_k are outside the constraint set. Starting from x_0, x_1, etc., we are outside the constraint set.

In the examples we considered, such as

minimizing $x_1 + x_2$

subject to

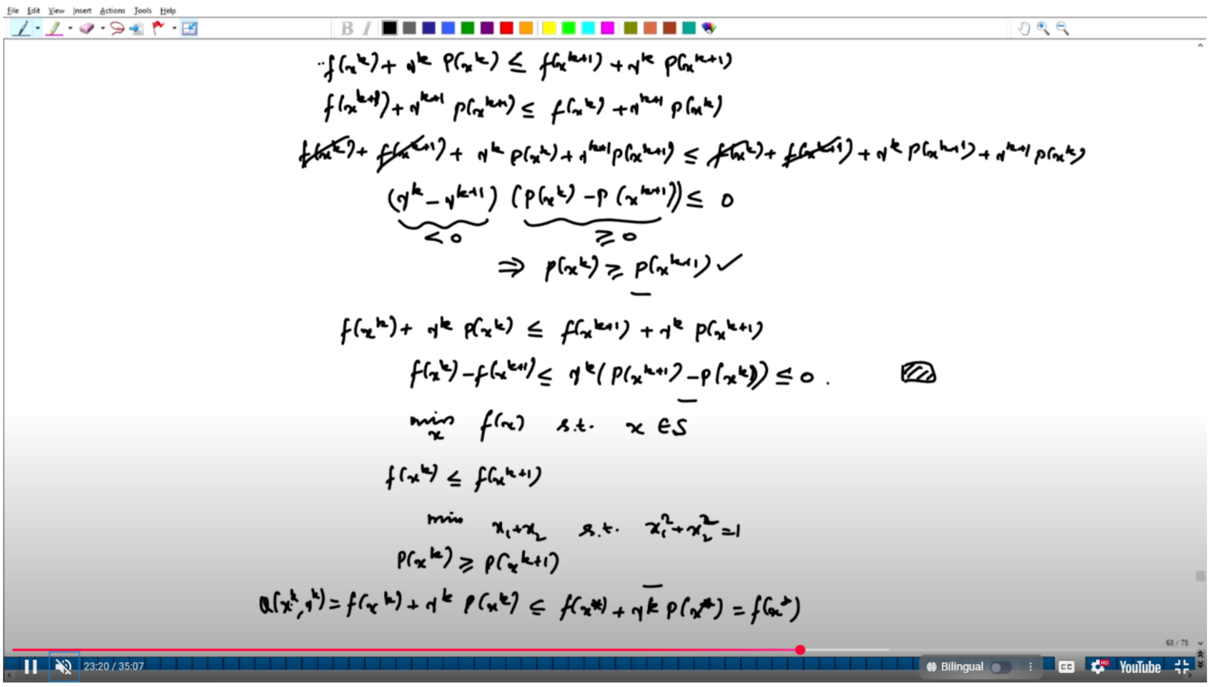
$x_1^2 + x_2^2 = 1$,

the initial points were outside the constraint set. The value of f increases as we move closer to the constraint set. The penalty p decreases, indicating that we are moving towards satisfying the constraints.

Note that for all k, $Q(x_k, \gamma_k) \leq f(x^*)$, where x^* is the actual minimizer subject to the constraints. Since $p(x^*) = 0$, we have $Q(x^*, \gamma_k) = f(x^*)$.

Therefore, $Q(x_k, \gamma_k) \leq f(x^*)$.

(Refer Slide Time 23:30)

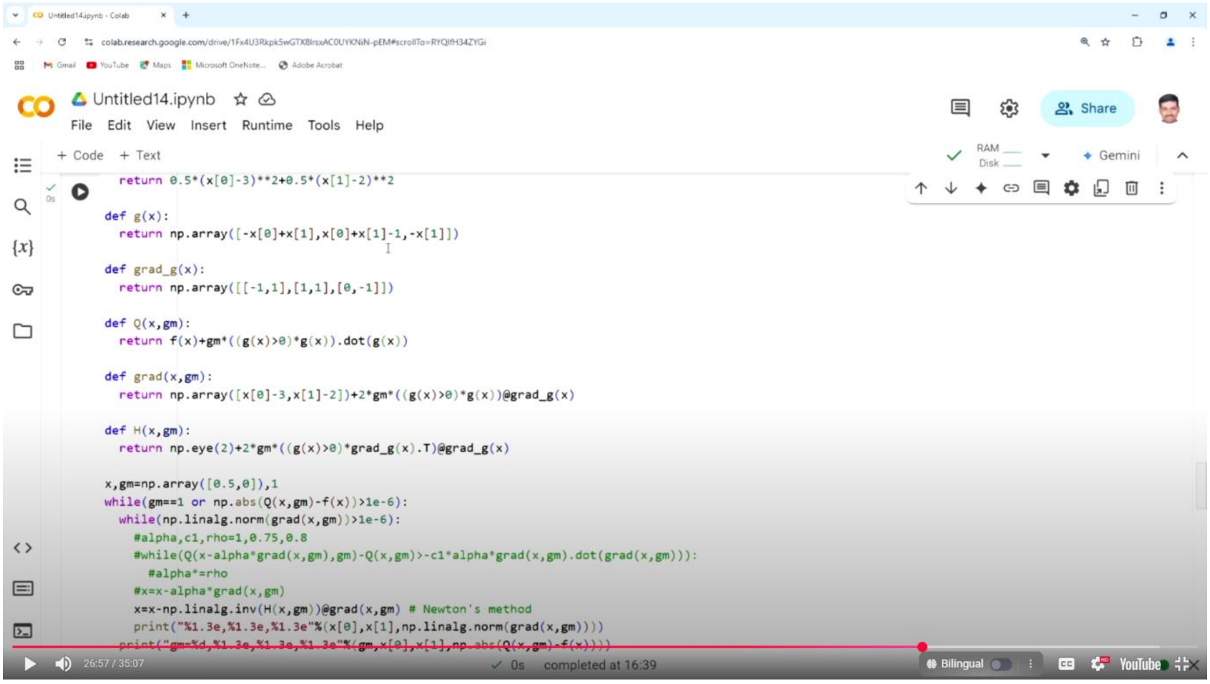


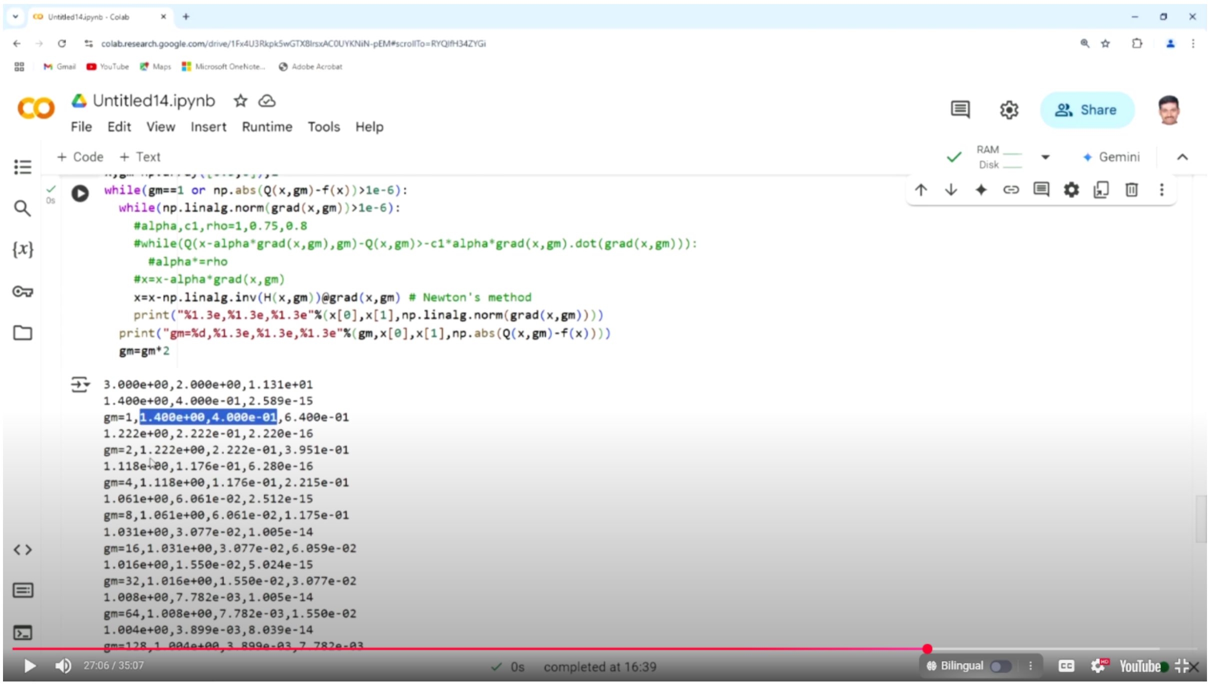
This means we are outside the constraint set and slowly moving towards $f(x^*)$. The method approaches the constraint set from outside.

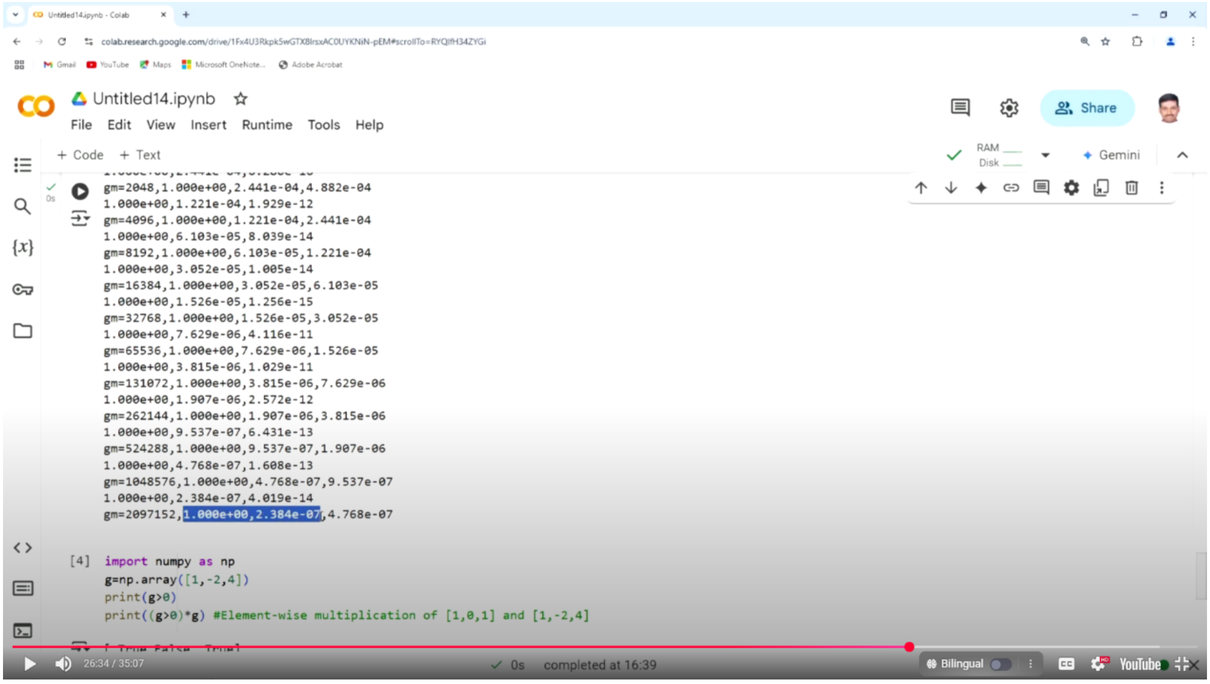
If the initial point is inside the constraint set, it will quickly move outside in the first iteration. For example, in the **problem with inequality constraints**

minimizing $\frac{1}{2}(x_1 - 3)^2 + \frac{1}{2}(x_2 - 2)^2$, if we start at (0.5, 0), which is inside the constraint set, the method immediately moves to points outside the constraint set, such as (1.4, 0.4), which violates $x_1 + x_2 \leq 1$. The method then converges to (1, 0).

(Refer Slides Time 26:00- 27:10)







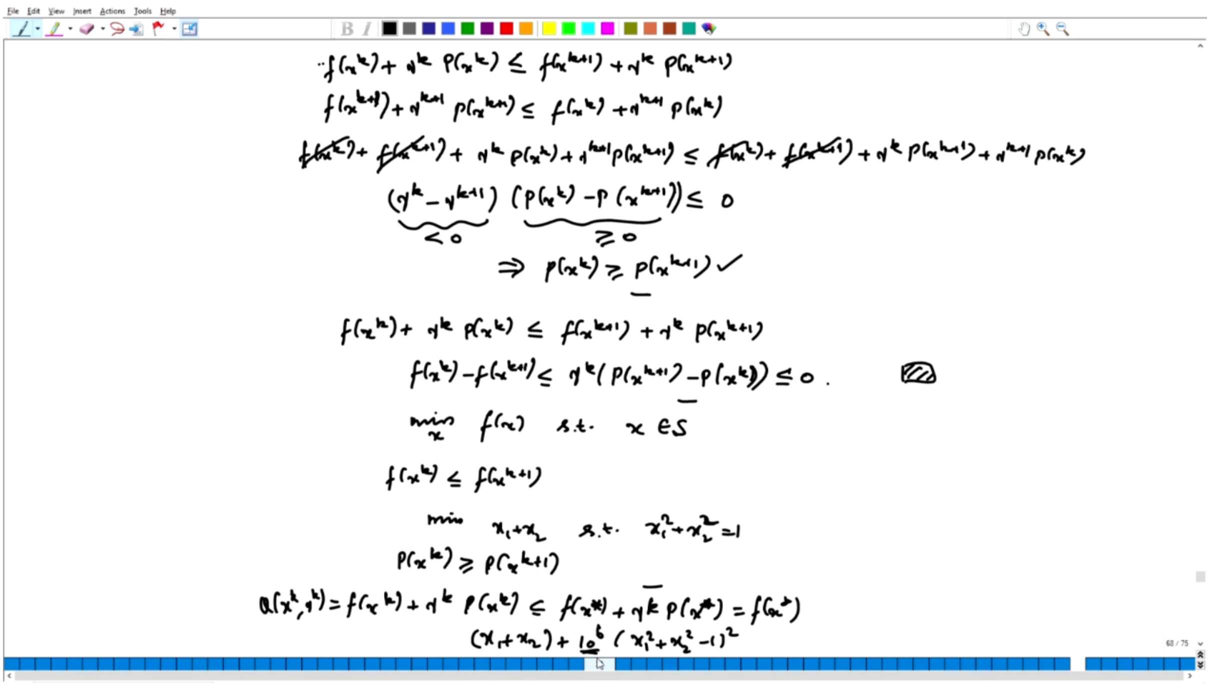
I will now point out two issues with the quadratic penalty method.

First, convergence occurs only for very high values of γ. You never converge for small values like $\gamma = 1$, 10, or 100. Typically, γ needs to be in the hundreds of thousands or millions. This is a problem because for high γ, the minimization becomes difficult. The condition number of the Hessian is far from 1, so methods like steepest descent require many iterations.

For example,

minimizing $x_1 + x_2 + 10^6 (x_1^2 + x_2^2 - 1)^2$ is challenging due to the large coefficient.

(Refer Slide Time 30:00)



Second, there can be issues where the minimization of $Q(x, \gamma)$ diverges for certain values of γ. Consider the problem:

minimize $x_1^2 - 15 x_2^2$

subject to

$x_2 = 2$.

The solution is $x_1 = 0$, $x_2 = 2$.

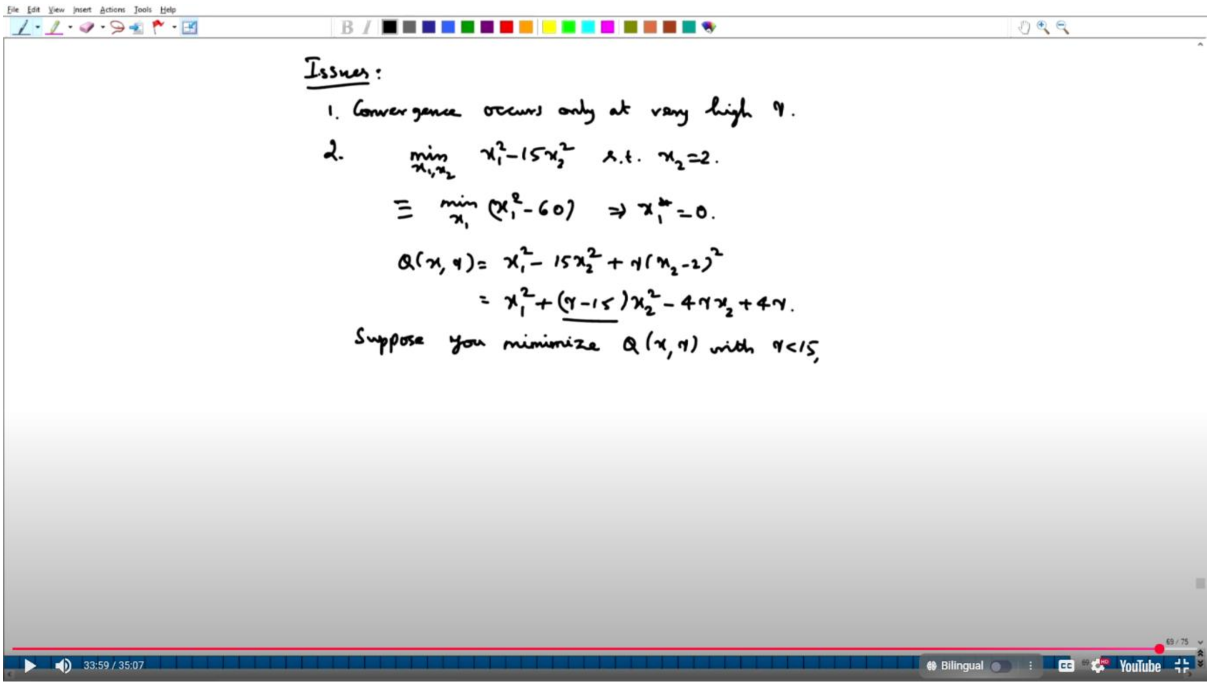
However, the quadratic penalty function is

$Q(x, \gamma) = x_1^2 - 15 x_2^2 + \gamma (x_2 - 2)^2 = x_1^2 + (\gamma - 15) x_2^2 - 4\gamma x_2 + 4\gamma$.

If $\gamma < 15$, then $\gamma - 15 < 0$, and $Q(x, \gamma)$ can go to $-\infty$ by taking $x_2 \to \infty$.

Thus, the minimization fails for $\gamma < 15$.

(Refer Slide Time 34:40)



These issues highlight limitations of the quadratic penalty method. For this reason, we will move to the augmented Lagrangian method, which we will cover next week.

This concludes our discussion of the quadratic penalty method.

Thank you.