Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematic

Institute of IIT Kanpur

Lecture: 36

Hello everyone, this is the first lecture of week 8. Recall that in week 7, we started with algorithms for constrained optimization and learned about the quadratic penalty method. We identified two major issues with the quadratic penalty method:

1. The penalty parameter γ must be extremely high, often in the millions, to achieve convergence, even for low-dimensional problems with few constraints.

2. For some problems, the minimum of the penalty function may not exist for certain values of γ, leading to divergence to negative infinity.

Due to these limitations, we now introduce a new method called the augmented Lagrangian method.

We begin by considering optimization problems with only equality constraints. The general form is:

Minimize $f(x)$ subject to $h_j(x) = 0$ for all $j = 1$ to m.

In the **quadratic penalty method**, we defined the penalty function as:

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

We then minimized $Q(x, \gamma)$ with respect to x for a fixed γ and increased γ iteratively.

In the **augmented Lagrangian method**, we define a different function:

$L_\gamma(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j h_j(x) + \gamma \sum_{j=1}^{m} [h_j(x)]^2$.

The key difference is the addition of the term $\sum_{j=1}^{m} \lambda_j h_j(x)$, which is the Lagrangian term from constrained optimization theory.

To understand why this is beneficial, recall that for the constrained problem min $f(x)$ subject to $h_j(x) = 0$, we form the Lagrangian:

$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j h_j(x)$.

The critical points satisfy:

$\nabla_x L(x, \lambda) = \nabla f(x) + \sum_{j=1}^{m} \lambda_j \nabla h_j(x) = 0$,

and $\nabla_\lambda L(x, \lambda)=0$ implies $h_j(x) = 0$ for all $j=1,2,..m$.

If the optimal Lagrange multipliers λ^* were known, then minimizing the Lagrangian function with respect to x would yield the critical points. Specifically,

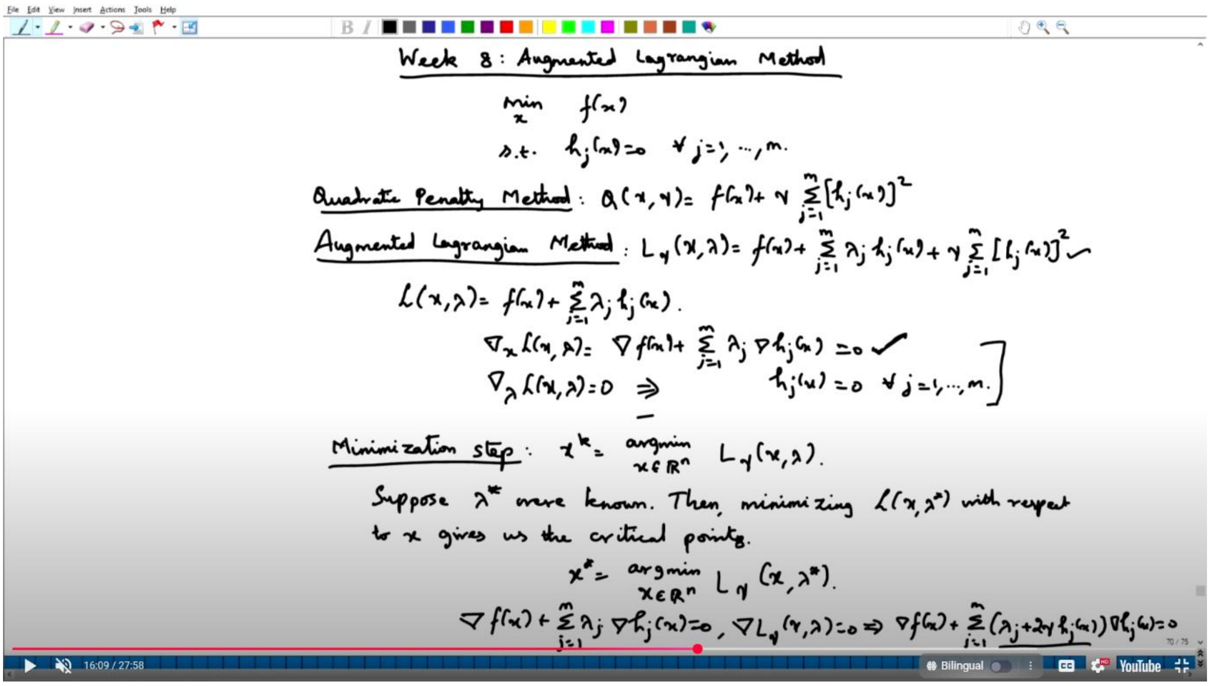
$x^* = \text{argmin}_{x \in \mathbb{R}^n} L(x, \lambda^*)$

However, since λ^* is unknown, we must estimate it iteratively. We begin with an initial guess λ_0 and update both x and λ in each iteration.

The update for λ comes from analyzing the optimality conditions. For the augmented Lagrangian:
$L\_\gamma(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j h_j(x) + \gamma \sum_{j=1}^{m} [h_j(x)]^2$

The gradient condition $\nabla_x L\_\gamma(x, \lambda) = 0$ gives: $\nabla f(x) + \sum_{j=1}^{m} (\lambda_j + 2\gamma h_j(x)) \nabla h_j(x) = 0$.

(Refer Slide Time 16:04)



This suggests that the effective Lagrange multipliers are $\lambda_j + 2\gamma h_j(x)$. Therefore, we update the Lagrange multipliers as: $\lambda_{k+1} = \lambda_k + 2\gamma h(x)$

Here, $h(x)$ is the vector of constraint values $[h_1(x), h_2(x), ..., h_m(x)]$.

The augmented Lagrangian method iteratively updates both x and λ. The algorithm is as follows:

1. Initialize x_0, λ_0, $\gamma > 0$, tolerance $\varepsilon > 0$, and set $k = 0$.

2. While $|L\_\gamma(x_k, \lambda_k) - f(x_k)| > \varepsilon$:

   a. Find $x_{k+1} = \arg\min_x L\_\gamma(x, \lambda_k)$ using an unconstrained optimization method (e.g., gradient descent, Newton's method).

   b. Update the Lagrange multipliers: $\lambda_{k+1} = \lambda_k + 2\gamma h(x_{k+1})$, where $h(x) = [h_1(x), ..., h_m(x)]^T$.

c. Set $k = k + 1$.

3. Output $x^* = x_k$ and $\lambda^* = \lambda_k$.

The quadratic penalty method and the augmented Lagrangian method share a similar algorithmic structure, but with key differences in initialization and update steps.

In the quadratic penalty method:

- Initialize a sequence of penalty parameters $\gamma_0, \gamma_1, \ldots$ that diverges to infinity
- Initialize x_0 and set a tolerance
- For each γ_k, minimize $Q(x, \gamma_k) = f(x) + \gamma_k \sum_{j=1}^{m} [h_j(x)]^2$ with respect to x
- Update $x_{k+1} = \text{argmin}_x Q(x, \gamma_k)$
- Check convergence using $h(x_{k+1})$

In the augmented Lagrangian method:

- Initialize a single penalty parameter $\gamma > 0$
- Initialize λ_0 and x_0
- For each iteration k:
  - Minimize $L\_\gamma(x, \lambda_k) = f(x) + \sum_{j=1}^{m} \lambda_j h_j(x) + \gamma \sum_{j=1}^{m} [h_j(x)]^2$ with respect to x
  - Update $\lambda_{k+1} = \lambda_k + 2\gamma h(x_{k+1})$
- Check convergence

The most significant advantages of the augmented Lagrangian method are:

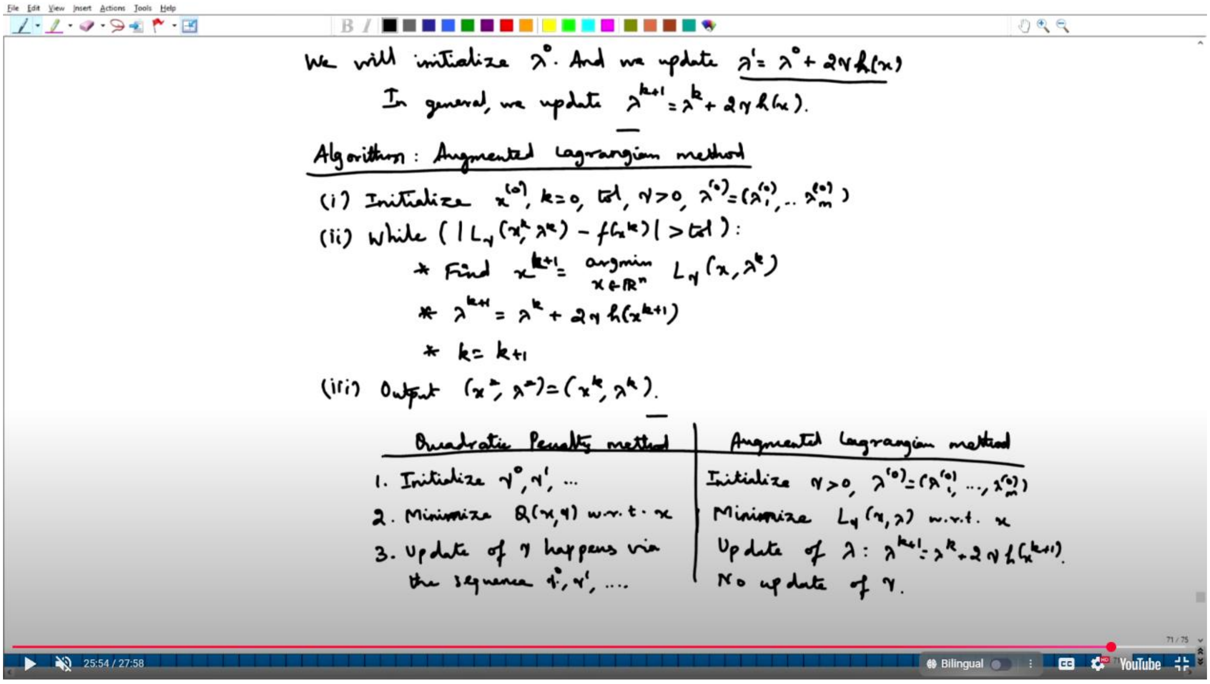
1. It works with moderate values of γ (e.g., $\gamma = 1$ or $\gamma = 10$) rather than requiring $\gamma \to \infty$
2. The Lagrange multipliers λ converge to their true optimal values
3. It avoids the numerical ill-conditioning issues that occur in the quadratic penalty method when γ becomes large
4. It prevents the solution divergence that can occur in the quadratic penalty method for certain problems

Additionally, the second issue mentioned at the end of the previous lecture (week 7, lecture 5) is resolved.

In that example, for certain positive values of γ, the minimum of the quadratic penalty function diverged to negative infinity, even though the true minimum existed at $(x_1^*, x_2^*) = (0, 2)$.

Such divergence problems do not occur with the augmented Lagrangian method.

(Refer Slide Time 27:50)



In the next lecture, we will illustrate these points with a concrete example focusing on equality constraints alone.

Thank you.