OM-S20-23: Towards More General Constrained Optimization

C. V. Jawahar

IIIT Hyderabad

http://preon.iiit.ac.in/om_quiz

14 Apr 2020

Announcements

- **Mid 1:** You might have seen the scores. If there are any specific issues left out on that, please email to Aditya Bharti. We shall try to address it in the best possible way.
- **Assignment:** Some of you had questions on Assignment 1 grades. A separate query handling session is scheduled.
- **Review Questions:** Some of you had questions. Most of them were possibly due to lack of dates along with that. See updated scores. A special query handling session is scheduled.
- **Home works:** Considering the internet issues and challenges with online mode, we are relaxing any time constraint on the questions. Shall loook into any server issues.

Announcements (Cont.)

- Any issues with the online mode, internet connection, please email *omonlinemodeclass@gmail.com*
- Please read instructions in team and or moodle. Please read the lecture slides/pdf files.
- Course goes until end of this month. We will be slowly winding up.
- Additional/Extra questions are there in HW, Reviews. Hope that takes care of special interruptions some of you have on specific days.
- We do NOT have any one with serious issue in continuing and completing the course online model. If you have or if you know, please inform us ASAP.

Announcements (Cont.)

- **ILM**You go through the videos, resources, sessions. You discuss, you solve problems, you answer questions, you discuss among yourselves.
- **ILM Participation** It measures how seriously you get involved. It need not be only the questions that we ask.
- **ILM Questions** There are 10 (simple) questions for you to answer. Posted (?) already for Manifold Learning. For the second ILM, it will be posted by next week.
- **Now** We are already into the last phase on “nonlinear optimization” with GD and Newtom’s methods.

Q01

'Gradient descent' often suffers from the problem of "getting stuck in local minima" while using in machine learning (say Deep learning) situations. Newton's method does not suffer with this limitation.
Do you agree with the above statement? Give a one sentence explanation.

Let us consider a machine learning problem where we want to optimize a function of the form

$$J(w) = \sum_{i=1}^{N} ||y_i - f(W, x_i)||_2^2$$

Consider a situation where we do not know how $f()$ is implemented (i.e., corresponding equation(s)). Let us assume it was a black box (say a chip that computes $f(W, x)$ given W and x).
How do we optimize $\min J(w)$ using gradient descent?
Why not we use the this method/idea in deep neural networks (they are often blamed to be backbox anyway)?Any disadvantage?

Q03

Why is backtracking not popular in gradient descent based optimization in Deep Learning (say backpropagation algorithm)?

Q04

Give five reasons why newtons method based optimization is not very popular in deep learning (training neural networks)?

Q05

We want to optimize

$$J(w) = \sum_{i=1}^{N} f(x_i, w)$$

Where w is a scalar quantity. $f()$ is a scalar valued function.
Write the update equation in gradient descent and using Newton's method.

Q06

We want to optimize

$$J(w) = \sum_{i=1}^{N} f(x_i, w)$$

Where $w \in R^p$ is a vector. $f()$ is a scalar valued function.
Write the update equation in gradient descent and using Newton's method.

Q07

We want to optimize

$$J(w) = \sum_{i=1}^{N} f(x_i, w)^T f(x_i, w)$$

Where $w \in R^p$ is a vector. $f()$ is a vector valued (R^q) function.
Write the update equation in gradient descent and using Newton's method.

Q08

Starting with an initialization closer to the minima, gradient descent based optimization converge (i) faster for convex optimization (ii) faster for non-convex optimization (iii) convergence does not depend on the problem nature. It depends only on the initialization.
(assume all other parameters constant for both)

Q09

We are solving $f(x) = 0$ with $f(x)$ being a quadratic function. We start with x^0 reasonably close to the root. We solve using newton's method. We reach the root (numerically close) in (i) one step (ii) two step (iii) many steps.
Agree/Disagree/Discuss in 1-2 sentences.

Q10

We are optimizing $f(x)$ with $f(x)$ being a quadratic function. We start with x^0 reasonably close to the optima. We optimize using newton's method.
We reach the optima (numerically close) in (i) one step (ii) two step (iii) many steps.
Agree/Disagree/Discuss in 1-2 sentences.

Constrained Optimization (Recap)

- We studies:

$$\max \; c^T x$$

Subject to:

$$Ax \leq b$$

$$x \geq 0$$

- We also Studied:

$$\max \; x^T A x$$

Subject to

$$||x|| = 1$$

Lagrange Multipliers

- We also Studied:

$$\max \ x^T A x$$

  Subject to

$$||x||_2^2 = 1$$

- Lagrangian:

$$L(x, \lambda) = x^T A x - \lambda(x^T x - 1)$$

$$Ax = \lambda x$$

Lagrange Multipliers

- Consider:

$$\max \ f(x)$$

  Subject to:

$$g(x) \leq b$$

- where $f(x)$ and $g(x)$ can be general nonlindear functions.
- We use Lagrange Multipliers:

$$L(x, \lambda) = f(x) - \lambda^T((g(x) - b)$$

Conditions for Optimality

- Unconstrained $f(x)$ $R \to R$
- Unconstrained $f(x)$ $R^n \to R$
- Comments on convex

Constrained Optimization

- Why is it not that simple?

Key Words

- Lagrangians
- KKT Conditions
- Lagrangian Duality
- Convex Optimization