OM-S20-21: Nonlinear Optimization

C. V. Jawahar

IIIT Hyderabad

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

03 Apr 2020

Problem Setting

- We are interested in solving two categories of problems:
  - Problem 1: Min $||g(x)||$
  - Problem 2: Solve $f(x) = 0$
  - f or g are nonlinear functions.
- In fact, both these problems are related. Second problem often arises in the optimization setting with derivative of the first equated to zero. i.e., a condition for optimality.
- A class of methods both used in
  - Convex optimization
  - Non-convex optimization
- The function f (or g) can be
  - $f : R \to R$
  - $f : R^n \to R$
  - $f : R^n \to R^m$
- f : convex $\to$ Local Minima $=$ Global Minima
- f : non-convex $\to$ Local Minima $\neq$ Global Minima

Terminology - I

Gradient: The gradient of a function $f(x)$ of n variables, at x^*, is the vector of first partial derivatives evaluated at x^*, and is denoted as $\nabla f(x^*)$:

$$\nabla f(x^*) = \begin{bmatrix} \frac{\partial f(x^*)}{\partial x_1} \\ \frac{\partial f(x^*)}{\partial x_2} \\ \vdots \\ \frac{\partial f(x^*)}{\partial x_n} \end{bmatrix} \tag{1}$$

Terminology - II

Hessian: The Hessain of a function $f(x)$ of n variables, at x^*, is the matrix of second partial derivatives evaluated at x^*, and is denoted as $\nabla^2 f(x^*)$,
Hessian is a square matrix of second-order partial derivatives of a function. It describes the local curvature of a function of many variables.

$$H = \nabla^2 f(x^*) = \begin{bmatrix} \frac{\partial^2 f(x^*)}{\partial x_1^2} & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x^*)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f(x^*)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x^*)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x^*)}{\partial x_n^2} \end{bmatrix} \tag{2}$$

Terminology - III

Jacobian: Given a set of m equations $y_i = f_i(x)$ in n variables $x_1, ..., x_n$, the Jacobian is defined as

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \frac{\partial f_m}{\partial x_2} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \tag{3}$$

Looking into the above definitions one can observe a simple relation between $\nabla f(x)$, $\nabla^2 f(x)$ and J.

The Jacobian of Gradient is Hessian.

i.e,

$$J(\nabla f(x)) = \nabla^2 f(x); \tag{4}$$

Solve $f(x) = 0$

- Iterative solutions.
    - Initialize
    - Refine
- Convergence
- Idea of fixed point iterations.
- Bisection
- Newton's method
- Secant method

On Bisection

If $c_1 = (a+b)/2$ is the midpoint of the initial interval, and c_n is the midpoint of the interval in the n th step, then the difference between c_n and the solution c^* is bounded by

$$|c_n - c^*| \leqslant \frac{|b-a|}{2^n}$$

This formula can be used to determine in advance the number of iterations that the bisection method would need to converge to a root to within a certain tolerance. The number of iterations needed, n, to achieve a given error (or tolerance), ϵ is given by

$$n = \log_2\left(\frac{\epsilon_0}{\epsilon}\right) = \frac{\log \epsilon_0 - \log \epsilon}{\log 2}$$

where, $\epsilon_0 = b - a$ (inital bracket size).

Fixed Point Iteration

The sequence defined by

$$p_n = f(p_{n-1}), n = 1, 2, \ldots, \infty$$

converges to the unique fixed point p^*, if $|f'(x)| \leqslant k < 1$ in the neighbourhood. The proof is straight-forward:

$$\begin{aligned} |p_n - p^*| &= |f(p_{n-1}) - f(p^*)| \\ &= \left|f'(r)\right| . |p_{n-1} - p^*| \, \text{by MVT} \\ &= k \left|p_{n-1} - p^*\right| \end{aligned}$$

Since $k < 1$, the distance to the fixed-point is shrinking in every iteration. In fact,

$$|p_{n-1} - p^*| \leqslant k^n |p_{n-1} - p^*|$$

Note: The **mean value theorem** tells $\exists r \in (p^*, q^*)$:

$$f'(r) = \frac{f(p^*) - f(q^*)}{p^* - q^*}$$

First order Approximation

$$f(y) = f(x) + \frac{f'(x)}{1!}(y-x) + \frac{f''(x)}{2!}(y-x)^2 + \frac{f^{(3)}(x)}{3!}(y-x)^3 + \cdots .$$

The first order approximation takes the first two terms in the series and approximates the function

$$f(y) = f(x) + \frac{f'(x)}{1!}(y-x)$$

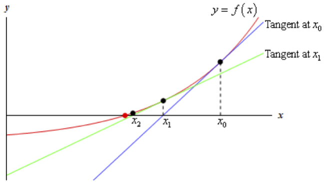
Let $f(y) = 0$, the true solution:

$$y - x = \frac{-f(x)}{f'(x)} \Rightarrow y = x - \frac{f(x)}{f'(x)}$$

$$x^{k+1} = \ x^k - f(x)/f'(x)$$

Newton's

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \text{ or } x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



If we are close to the root, then $|x - x^*|$ is small, which means that $|x - x^*|^2 \ll |x - x^*|$, hence we make the approximation:

$$0 \approx f(x) + (x^* - x)\, f'(x), \leftrightarrow x^* \approx x - \frac{f(x)}{f'(x)}$$

Newton's and Secant

Newton's method as a fixed-point iteration: $x_n = x_{n-1} - \frac{f(x_{n-1})}{f'(x_{n-1})}$.
Then (the fixed point theorem), we must find an interval $[x^* - \delta, x^* + \delta]$ that g maps into itself, and for which $|g'(x)| \leqslant k < 1$.

$$g'(x) = 1 - \frac{f'(x)f'(x) - f(x)f''(x)}{[f'(x)]^2} = \frac{f(x)f''(x)}{[f'(x)]^2}$$

By assumption $f(x^*) = 0, f'(x^*) \neq 0$, so $g'(x^*) = 0$. By continuity $|g'(x)| \leqslant k < 1$ for some neighborhood of x^*. Hence the fixed-point iteration will converge.

Secant:

$$x^{k+1} = x^k - f(x^k)\frac{x^k - x^{k-1}}{f(x^k) - f(x^{k-1})}$$

First order Approximation

$$f(y) = f(x) + \frac{f'(x)}{1!}(y-x) + \frac{f''(x)}{2!}(y-x)^2 + \frac{f^{(3)}(x)}{3!}(y-x)^3 + \cdots .$$

The first order approximation takes the first two terms in the series and approximates the function

$$f(y) = f(x) + \frac{f'(x)}{1!}(y-x)$$

Let $f(y) = 0$, the true solution:

$$y - x = \frac{-f(x)}{f'(x)} \Rightarrow y = x - \frac{f(x)}{f'(x)}$$

$$x^{k+1} = \; x^k - f(x)/f'(x)$$

$f: \mathbf{R^n} \rightarrow \mathbf{R^m}$

$$\overline{f}(y) = \overline{f}(x) + J_x(y - x)$$

Where Jacobian J is an $m \times n$ matrix and can be given as

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Putting $f(y) = 0$ in equation we get:

$$\overline{0} = \overline{f}(\overline{x}) + J_x(\overline{y} - \overline{x})$$

Let $\overline{y} - \overline{x} = \overline{s}$

$$0 = f(x) + J_x s$$

$$J_x s = -f(x)$$

Procedure

$\overline{x}^0 \leftarrow$ Initial guess
$k = 0, 1, 2, \cdots$ do
Solve $J_{x_k} s = -f(x_k)$ for s
$x_{k+1} = x_k + s$

Next Steps

- Start going through the videos
- Discuss online
- More in the next lecture session to follow the videos