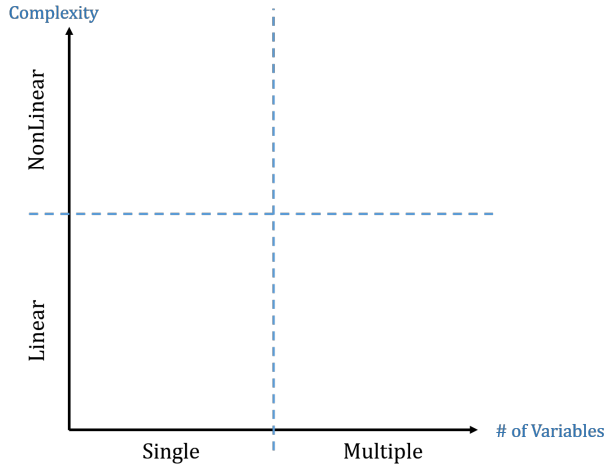


Class Notes: Applied Optimization Techniques

Muhammad Obaidullah

M.A.Sc. Candidate, Ryerson University, mobaidullah@ryerson.ca

I. LECTURE 2: 7th OCTOBER 2015



A. Newton-Raphson Technique

1) Single Variable:

$$x_{new} = x_{old} - \frac{g(x_{old})}{g'(x_{old})} \quad (1)$$

Example: Single Variable

Let's take the following equation

$$f(x) = x^4 + 3x^3 + 5x^2 - 2x + 8 \quad (2)$$

$+\infty$ and $-\infty$ are always the roots of the polynomials. Let's not consider that. Let's be realistic.

$$g(x) = 4x^3 + 9x^2 + 10x - 2 = 0 \quad (3)$$

$$x_{new} = x_{old} - \frac{g(x_{old})}{g'(x_{old})} \quad (4)$$

$$= x_{old} - \frac{4x_{old}^3 + 9x_{old}^2 + 10x_{old} - 2}{12x_{old}^2 + 18x_{old} + 10} \quad (5)$$

If $x_{old} = 0$

$$x_{new} = 0 - \frac{-2}{+10} = \frac{2}{10} \quad (6)$$

If $x_{old} = \frac{2}{10}$

$$x_{new} = \frac{2}{10} - \frac{4(0.2)^3 + 9(0.2)^2 + 10(0.2) - 2}{12(0.2)^2 + 18(0.2) + 10} \quad (7)$$

$$= \frac{2}{10} - \frac{0.392}{14.08} \quad (8)$$

$$= 0.17 \quad (9)$$

So these are the two iteration of this technique. In exam only two iterations will be asked.

2) Multiple Variable:

$$\begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix}$$

Example: Multiple Variable

$$f(x, y) = x^3 + 3xy^2 + 4x^2y + 5y^3 - 6x - 8y \quad (10)$$

$$\frac{\partial f}{\partial x} = g_1(x, y) = 0 = 3x^2 + 3y^2 + 8xy - 6 \quad (11)$$

$$\frac{\partial f}{\partial y} = g_2(x, y) = 0 = 6xy + 4x^2 + 15y^2 - 8 \quad (12)$$

$$\begin{bmatrix} \frac{\partial g_1}{\partial x} & \frac{\partial g_1}{\partial y} \\ \frac{\partial g_2}{\partial x} & \frac{\partial g_2}{\partial y} \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = \begin{bmatrix} g_1(x, y) \\ g_2(x, y) \end{bmatrix}$$

$$\begin{bmatrix} 6x + 8y & 6y + 8y \\ 6y + 8x & 6x + 30y \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = - \begin{bmatrix} 3x^2 + 3y^2 + 8xy - 6 \\ 6xy + 4x^2 + 15y^2 - 8 \end{bmatrix}$$

1st Iteration Let $x_{old} = 1$ and $y_{old} = 1$, then:

$$\begin{bmatrix} 6 + 8 & 6 + 8 \\ 6 + 8 & 6 + 30 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = - \begin{bmatrix} 3 + 3 + 8 - 6 \\ 6 + 4 + 15 - 8 \end{bmatrix}$$

Therefore,

$$\Delta x = -0.9 \quad (13)$$

$$\Delta y = \frac{9}{22} = -0.375 \quad (14)$$

$$x_{new} = x_{old} + \Delta x = 1 + (-0.9) = 0.1 \quad (15)$$

$$y_{new} = y_{old} + \Delta y = 1 + (0.375) = 1.375 \quad (16)$$

Checking convergence by Euclidean distance

$$d_{Euclidean} = \sqrt{(1 - 0.1)^2 + (1 - 0.1375)^2} \leq 0.02 \quad (17)$$

Example: Multiple Variable

$$z = y = x = 2x^2 - 2xy - y^2 \quad (18)$$

$$\frac{\partial z}{\partial x} = -1 - 4x - 2y = 0 = g_1 \quad (19)$$

$$\frac{\partial z}{\partial y} = 1 - 2x - 2y = 0 = g_2 \quad (20)$$

$$\begin{bmatrix} -4 & -2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = - \begin{bmatrix} -1 - 4x - 2y \\ 1 - 2x - 2y \end{bmatrix}$$

Let's begin with $x_{old} = 0$ and $y_{old} = 0$

$$\begin{bmatrix} -4 & -2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} = - \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

$$\Delta x = -1 \text{ and } \Delta y = \frac{3}{2}$$

$$\begin{bmatrix} -4 & -2 \\ -2 & -2 \end{bmatrix} \begin{bmatrix} -1 \\ \frac{3}{2} \end{bmatrix} = \begin{bmatrix} 4 - 3 \\ 2 - 3 \end{bmatrix}$$

$$x_{new} = 0 + (-1) \quad (21)$$

$$y_{new} = 0 + \left(\frac{3}{2}\right) \quad (22)$$

B. Linear Systems

1) *LU Decomposition*: Lower triangular matrix is easily to handle because all variable an be calculated by forward operation

$$\begin{bmatrix} a_1 & 0 & 0 \\ a_2 & b_2 & 0 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

Upper triangular matrix is easily to handle because all variable an be calculated by backward operation

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ 0 & b_2 & c_2 \\ 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix}$$

If you have

$$A X = K \quad (23)$$

it can easily be broken down into LU components. (Lower and Upper matrix components)

For example if you have following system of equations:-

$$\begin{cases} a_1x + b_1y + c_1z + d_1\theta = K_1 \\ a_2x + b_2y + c_2z + d_2\theta = K_2 \\ a_3x + b_3y + c_3z + d_3\theta = K_3 \\ a_4x + b_4y + c_4z + d_4\theta = K_4 \end{cases}$$

it can easily be broken down into LU components. (Lower and Upper matrix components) as follows

$$\begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & u_{13} & u_{14} \\ 0 & u_{22} & u_{23} & u_{24} \\ 0 & 0 & u_{33} & u_{34} \\ 0 & 0 & 0 & u_{44} \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ z \\ \theta \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ K_4 \end{bmatrix}$$

The upper triangular matrix can be consumed by the unknown single column matrix and results in:-

$$\begin{bmatrix} l_{11} & 0 & 0 & 0 \\ l_{21} & l_{22} & 0 & 0 \\ l_{31} & l_{32} & l_{33} & 0 \\ l_{41} & l_{42} & l_{43} & l_{44} \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ \zeta \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ K_4 \end{bmatrix}$$

So this allows the linear system of equation to be broken into lower and upper matrices calculated once only and then can be used by the program to calculate faster and accurate results later.

Example

You want to decompose the following matrix:

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{bmatrix}$$

$$\left[\begin{array}{cccc} a'_{11} & a'_{12} = \frac{a_{12}}{a'_{11}} & a'_{13} = \frac{a_{13}}{a'_{11}} & a'_{14} = \frac{a_{14}}{a'_{11}} \\ a'_{21} & a'_{22} = a_{22} - a'_{12}a'_{21} & a'_{23} = \frac{a_{23} - a_{13}a'_{21}}{a'_{22}} & a'_{24} = \frac{a_{24} - a_{14}a'_{21}}{a'_{22}} \\ a'_{31} & a'_{32} = a_{32} - a'_{12}a'_{31} & a'_{33} = a_{33} - (a'_{13}a'_{31} + a'_{23}a'_{32}) & a'_{34} = \frac{a_{34} - (a'_{14}a'_{31} + a'_{24}a'_{32})}{a'_{33}} \\ a'_{41} & a'_{42} = a_{42} - a'_{12}a'_{41} & a'_{43} = a_{43} - (a'_{13}a'_{41} + a'_{23}a'_{41}) & a'_{44} = a_{44} - (a'_{14}a'_{41} + a'_{24}a'_{42} + a'_{34}a'_{43}) \end{array} \right]$$

This matrix contains both, the upper triangular matrix and the lower triangular matrix

C. Matrix Property

A matrix is positive (definite) if

$$Y' \mathbb{A} Y = > 0 \text{ for all } Y \in R^N \quad (24)$$

A matrix is negative (definite) if

$$Y' \mathbb{A} Y = < 0 \text{ for all } Y \in R^N \quad (25)$$

$$A = \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \end{bmatrix} \quad A = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$$

$$[x \ y] \begin{bmatrix} 3 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} > 0$$

$$[x \ y] \begin{bmatrix} 3x+2y \\ 2x+4y \end{bmatrix} = 3x^2 + 2xy + 2xy + y^2 > 0$$

$$2x^2 + (x+2y)^2 \geq 0 \quad (26)$$

So this is definite positive matrix

Now if

$$A = \begin{bmatrix} 5 & 7 \\ -2 & -9 \end{bmatrix}$$

then

$$[x \ y] \begin{bmatrix} 5 & 7 \\ -2 & -9 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = [x \ y]$$

$$\begin{bmatrix} 5x+7y \\ -2x-9y \end{bmatrix} = 5x^2 + 7xy - 2xy - 9y^2 = 5x^2 + 5xy - 9y^2$$

This is a saddle point

II. LECTURE 3: 21st OCTOBER 2015

A. Agenda

- 1) Linear Systems (*Continued*)
- 2) Optimization by Approximation (Curve fitting)
- 3) Gradient techniques
 - Steepest
 - a) Ascent
 - i) Fixed
 - ii) Optimum
 - b) Descent
 - i) Fixed
 - ii) Optimum

B. Linear Systems

$$\begin{cases} a_1x + b_1y + c_1z = d_1 \\ a_1x + b_2y + c_2z = d_2 \\ a_3x + b_3y + c_3z = d_3 \end{cases}$$

$$\begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \\ d_3 \end{bmatrix} \Rightarrow LUX = D$$

Cramer's Rule

$$AX = B \quad (27)$$

$$x = \frac{|A_x|}{|A|} \quad (28)$$

$$y = \frac{|A_y|}{|A|} \quad (29)$$

$$z = \frac{|A_z|}{|A|} \quad (30)$$

where

$$A_x = \begin{bmatrix} d_1 & b_1 & c_1 \\ d_2 & b_2 & c_2 \\ d_3 & b_3 & c_3 \end{bmatrix} \quad A_y = \begin{bmatrix} a_1 & d_1 & c_1 \\ a_2 & d_2 & c_2 \\ a_3 & d_3 & c_3 \end{bmatrix} \quad A_z = \begin{bmatrix} a_1 & b_1 & d_1 \\ a_2 & b_2 & d_2 \\ a_3 & b_3 & d_3 \end{bmatrix}$$

$$\begin{bmatrix} a_1 & b_1 \\ c_1 & d_1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \frac{1}{\Delta} \begin{bmatrix} d_1 & -b_1 \\ -c_1 & a_1 \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$$

Gauss-Sidel

$$x = \frac{d_1 - b_1y - c_1z}{a_1} \quad (31)$$

$$y = \frac{d_2 - a_2x - c_2z}{b_2} \quad (32)$$

$$z = \frac{d_3 - a_3x - b_2y}{c_3} \quad (33)$$

LU with pivoting

Gauss-Jarobi

C. Optimization by approximation (Curve fitting)

We are going to use quadratic curve fitting. In other words, we are going to fit any curve into a quadratic approximation curve.

$$y = ax^2 + bx + c \quad (34)$$

$$y' = 2ax + b = 0 \quad (35)$$

$$\therefore x_{opt.} = \frac{-b}{2a} \quad (36)$$

$$y_{opt.} = a(x_{opt.})^2 + b(x_{opt.}) + c \quad (37)$$

Now any three points $(x_1, y_1), (x_2, y_2), (x_3, y_3)$ can be taken from your graph and then the following system of equations should be solved to find $a, b, \text{ and } c$:

$$\begin{cases} y_1 = ax_1^2 + bx_1 + c \\ y_2 = ax_2^2 + bx_2 + c \\ y_3 = ax_3^2 + bx_3 + c \end{cases}$$

D. Steepest Ascent - Fixed

This is used when we are looking for a maximum of a function and is used with a constant/fixed multiple α . α determines how big or small the step is taken towards the optimum point. So choosing a right value for α is crucial so that the optimization algorithm does not overshoot or explode.

$$x_{new} = x_{old} + \alpha \frac{\partial f}{\partial x} \quad (38)$$

$$y_{new} = y_{old} + \alpha \frac{\partial f}{\partial y} \quad (39)$$

E. Steepest Descent - Fixed

This is used when we are looking for a minimum of a function and is used with a constant/fixed multiple α . α determines how big or small the step is taken towards the optimum point. So choosing a right value for α is crucial so that the optimization algorithm does not overshoot or explode.

$$x_{new} = x_{old} - \alpha \frac{\partial f}{\partial x} \quad (40)$$

$$y_{new} = y_{old} - \alpha \frac{\partial f}{\partial y} \quad (41)$$