Numerical Methods

DS288 and UMC201

Ratikanta Behera

Department of Computational and Data Sciences,
Indian Institute of Science Bangalore

August-December 2025

Newton's Method

$$\begin{cases} f_1(x_1, x_2) = 0 \\ f_2(x_1, x_2) = 0 \end{cases} \tag{1}$$

- Supposing that (x_1, x_2) is an approximate solution of (1),
- let us compute corrections h_1 and h_2 so that $(x_1 + h_1, x_2 + h_2)$ will be a better approximate solution.
- Using only linear terms in the Taylor expansion in two variables, we have

$$\begin{cases} 0 = f_1(x_1 + h_1, x_2 + h_2) \approx f_1(x_1, x_2) + h_1 \frac{\partial f_1}{\partial x_1} + h_2 \frac{\partial f_1}{\partial x_2} \\ 0 = f_2(x_1 + h_1, x_2 + h_2) \approx f_2(x_1, x_2) + h_1 \frac{\partial f_2}{\partial x_1} + h_2 \frac{\partial f_2}{\partial x_2} \end{cases} \tag{2}$$

- The partial derivatives appearing in (2) are to be evaluated at (x_1, x_2). Equation (2) constitutes a pair of *linear* equations for determining h_1 and h_2. The coefficient matrix is the **Jacobian matrix** of f_1 and f_2:

$$J = \begin{bmatrix} \partial f_1/\partial x_1 & \partial f_1/\partial x_2 \\ \partial f_2/\partial x_1 & \partial f_2/\partial x_2 \end{bmatrix}$$

Newton's Method

- To solve (2), we require J to be nonsingular. The solution is

$$\begin{bmatrix} h_1 \\ h_2 \end{bmatrix} = -J^{-1} \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}$$

- Hence, Newton's method for two nonlinear equations in two variables is

$$\begin{bmatrix} x_1^{(k+1)} \\ x_2^{(k+1)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} + \begin{bmatrix} h_1^{(k)} \\ h_2^{(k)} \end{bmatrix} = \begin{bmatrix} x_1^{(k)} \\ x_2^{(k)} \end{bmatrix} - J^{-1} \begin{bmatrix} f_1(x_1, x_2) \\ f_2(x_1, x_2) \end{bmatrix}, \quad k = 0, 1 \cdots$$

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - J(\mathbf{x}^{(k)})^{-1}\mathbf{F}(\mathbf{x}^{(k)}), \quad \text{for } k = 0, 1 \cdots .$$

Note: Solving the Jacobian system may be difficult if J is nearly singular.

Newton's Method

- The convergence of the method depends on initial approximation $\mathbf{x}^{(0)}$.
- A sufficient condition for convergence is that for each k

$$\|J^k\| < 1 \tag{3}$$

- However a necessary and sufficient condition for convergence is

$$\rho(J^k) < 1 \tag{4}$$

- Here $||.||$ is suitable norm and $\rho(J^k)$ is the spectral radius (largest eigenvalue in magnitude) of the matrix J^k.
- If the method converges, then its rate of convergence is two. The iterations are stopped when

$$\|x^{(k+1)} - x^{(k)}\| \leq \varepsilon \tag{5}$$

Newton's Method example

Example

The exact solution is $x = 2$ and $y = 1$ Compute three iterations of the Newton's method to solve the system of equations with the initial approximation as $x_0 = 1.5$, and $y_0 = 0.5$.

$$x^2 + xy + y^2 = 7 \tag{6}$$
$$x^3 + y^3 = 9 \tag{7}$$

$$f(x, y) = x^2 + xy + y^2 - 7 \tag{8}$$
$$g(x, y) = x^3 + y^3 - 9 \tag{9}$$

$$J_k = \begin{bmatrix} f_x(x_k, y_k) & f_y(x_k, y_k) \\ g_x(x_k, y_k) & g_y(x_k, y_k) \end{bmatrix} = \begin{bmatrix} 2x_k + y_k & x_k + 2y_k \\ 3x_k^2 & 3y_k^2 \end{bmatrix} \tag{10}$$

$$J_k^{-1} = \frac{1}{D_k} \begin{bmatrix} 3y_k^2 & -(x_k + 2y_k) \\ -3x_k^2 & 2x_k + y_k \end{bmatrix} \tag{11}$$

where, $D_k = |J_k| = 3y_k^2(2x_k + y_k) - 3x_k^2(x_k + 2y_k)$.

Newton's Method example

We can now write the method as

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} - J(\mathbf{x}^{(k)})^{-1}\mathbf{F}(\mathbf{x}^{(k)}), \quad \text{for } k = 0, 1 \cdots .$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \frac{1}{D_k} \begin{bmatrix} 3y_k^2 & -(x_k + 2y_k) \\ -3x_k^2 & 2x_k + y_k \end{bmatrix} \begin{bmatrix} x_k^2 + x_k y_k + y_k^2 - 7 \\ x_k^3 + y_k^3 - 9 \end{bmatrix} \tag{12}$$

where $k = 0, 1, \cdots$. Using $(x_0, y_0) = (1.5, 0.5)$, we get

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 1.5 \\ 0.5 \end{bmatrix} - \frac{1}{-14.25} \begin{bmatrix} 0.75 & -2.5 \\ -6.75 & 3.5 \end{bmatrix} \begin{bmatrix} -3.75 \\ -5.50 \end{bmatrix} = \begin{bmatrix} 2.2675 \\ 0.9254 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.2675 \\ 0.9254 \end{bmatrix} - \frac{1}{-49.4951} \begin{bmatrix} 2.5691 & -4.1183 \\ -15.4247 & 5.4604 \end{bmatrix} \begin{bmatrix} 1.0963 \\ 3.4510 \end{bmatrix} = \begin{bmatrix} 2.0373 \\ 0.9645 \end{bmatrix}$$

$$\begin{bmatrix} x_3 \\ y_3 \end{bmatrix} = \begin{bmatrix} 2.0373 \\ 0.9645 \end{bmatrix} - \frac{1}{-35.3244} \begin{bmatrix} 2.7908 & -3.9663 \\ -12.4518 & 5.0391 \end{bmatrix} \begin{bmatrix} 0.0458 \\ 0.3532 \end{bmatrix} = \begin{bmatrix} 2.0013 \\ 0.9987 \end{bmatrix}$$

Iteration method

Consider the following system:

$$f(x,y) = 0 \tag{13}$$

$$g(x,y) = 0 \tag{14}$$

Write may write this system in an equivalent form as

$$x = F(x,y) \tag{15}$$

$$y = G(x,y) \tag{16}$$

Let (ζ, η) be the solution. Therefore, (ζ, η) satisfies the equations

$$\zeta = F(\zeta, \eta) \tag{17}$$

$$\eta = G(\zeta, \eta) \tag{18}$$

Iteration method

- Let (x_0, y_0) be the suitable approximation to (ζ, η). Then we write a general iteration method

$$
\begin{aligned}
x_{k+1} &= F(x_k, y_k) && (19)\\
y_{k+1} &= G(x_k, y_k) \quad k = 0, 1, 2, \cdots && (20)
\end{aligned}
$$

- If the method converges, then $\lim\limits_{k\to\infty} x_k = \zeta$ and $\lim\limits_{k\to\infty} y_k = \eta$.
- Following the above equations we can write

$$
\begin{aligned}
\zeta - x_{k+1} = F(\zeta, \eta) - F(x_k, y_k) && (21)\\
\eta - y_{k+1} = G(\zeta, \eta) - G(x_k, y_k) && (22)
\end{aligned}
$$

- Let $\epsilon_k = \zeta - x_k$ and $\delta_k = \eta - y_k$ be the errors in the kth iteration. Now we can write

$$
\begin{aligned}
\epsilon_{k+1} &= F(x_k + \epsilon_k, y_k + \delta_k) - F(x_k, y_k) && (23)\\
\delta_{k+1} &= G(x_k + \epsilon_k, y_k + \delta_k) - G(x_k, y_k) && (24)
\end{aligned}
$$

Iteration method(special case)

- Expanding in Taylor series about (x_k, y_k) and neglecting the second and higher power of ϵ_k, δ_k, we obtain

$$\begin{aligned}
\epsilon_{k+1} &= \epsilon_k F_x(x_k, y_k) - \delta_k F_y(x_k, y_k) \\
\delta_{k+1} &= \epsilon_k G_x(x_k, y_k) - \delta_k G_y(x_k, y_k)
\end{aligned}$$

- This can be written as $$\epsilon^{k+1} = \mathbf{A}_k \epsilon^k \tag{25}$$
- where $\epsilon^{(k)} = [\epsilon_k, \delta_k]^T$ and $A_k = \begin{pmatrix} F_x & F_y \\ G_x & G_y \end{pmatrix}$
- Here A_k is the Jacobian matrix of the iteration functions F and G evaluated at (x_k, y_k).
- The sufficient condition for convergence is $||A_k|| < 1$ for each k

$$\begin{aligned}
|F_x(x_k, y_k)| + |F_y(x_k, y_k)| < 1 \\
|G_x(x_k, y_k)| + |G_y(x_k, y_k)| < 1
\end{aligned}$$

Iteration method

- The necessary and sufficient condition for convergence is that for each k

$$\rho(\mathbf{A_k}) < 1 \tag{26}$$

- where $\rho(A_k)$ is the spectral radius of the matrix A_k.
- If (x_0, y_0) is a close approximation to the root (ζ, η), then we usually test the conditions $||A_k|| < 1$ at the initial approximation (x_0, y_0).

Iteration method

Example

The system of equations

$$\begin{aligned} f(x, y) &= x^2 + 3x + y - 5 = 0 \\ g(x, y) &= x^2 + 3y^2 - 4 = 0 \end{aligned}$$

has a solution $(1, 1)$. Determine the iteration functions $F(x, y)$ and $G(x, y)$ so that the sequence of iterations obtained from

$$x_{k+1} = F(x_k, y_k) \tag{27}$$

$$y_{k+1} = G(x_k, y_k) \quad k = 0, 1, 2, \cdots \tag{28}$$

$(x_0, y_0) = (0.5, 0.5)$ converges to the root. Find first five iterations.

Frame Title

We write the given system of equations in an equivalent form

$$x = x + \alpha(x^2 + 3x + y - 5) = F(x, y) \tag{29}$$
$$y = y + \beta(x^2 + 3y^2 - 4) = G(x, y) \tag{30}$$

where α and β are arbitrary parameters, which are to be determined. If we use the maximum absolute row sum norm, we require that

$$|F_x(x_0, y_0)| + |F_y(x_0, y_0)| < 1 \tag{31}$$

and

$$|G_x(x_0, y_0)| + |G_y(x_0, y_0)| < 1. \tag{32}$$

Differentiating F and G partially with respect to x and y and evaluating at $(x_0, y_0) = (0.5, 0.5)$, we get

$$F_x(x, y) = 1 + (2x + 3)\alpha, \qquad F_x(0.5, 0.5) = 1 + 4\alpha \tag{33}$$
$$F_y(x, y) = \alpha, \qquad F_y(0.5, 0.5) = \alpha \tag{34}$$
$$G_x(x, y) = 2\beta x, \qquad G_x(0.5, 0.5) = \beta \tag{35}$$
$$G_y(x, y) = 1 + 6\beta y, \qquad G_y(0.5, 0.5) = 1 + 3\beta \tag{36}$$

Frame Title

Therefore, the conditions of convergence become

$$\begin{aligned} |1+4\alpha| + |\alpha| &< 1 \\ |\beta| + |1+3\beta| &< 1. \end{aligned} \tag{37}$$

Any values of α, β which satisfy (37) can be used. Obviously, both α and β are negative. Taking $\alpha = -1/4$ and $\beta = -1/6$, we obtain the iteration method

$$x_{k+1} = x_k - \frac{1}{4}(x_k^2 + 3x_k + y_k - 5) = -\frac{1}{4}(x_k^2 - x_k + y_k - 5) = F(x_k, y_k)$$
$$y_{k+1} = y_k - \frac{1}{6}(x_k^2 + 3y_k^2 - 4) = -\frac{1}{6}(x_k^2 + 3y_k^2 - 6y_k - 4) = G(x_k, y_k).$$

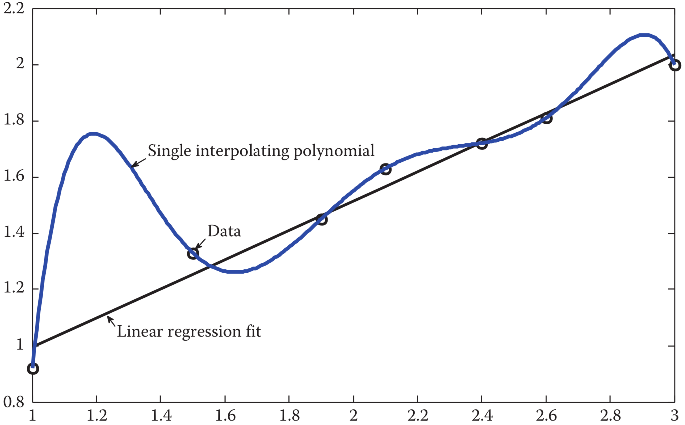
Starting with $(x_0, y_0) = (0.5, 0.5)$, we get

$$\begin{aligned} (x_1, y_1) &= (1.1875, 1.0), & (x_2, y_2) &= (0.944336, 0.931641), \\ (x_3, y_3) &= (1.030231, 1.015702), & (x_4, y_4) &= (0.988288, 0.989647), \\ (x_5, y_5) &= (1.005482, 1.003828). \end{aligned}$$

Chapter-8
Least Squares Approximation

Least Squares Approximation

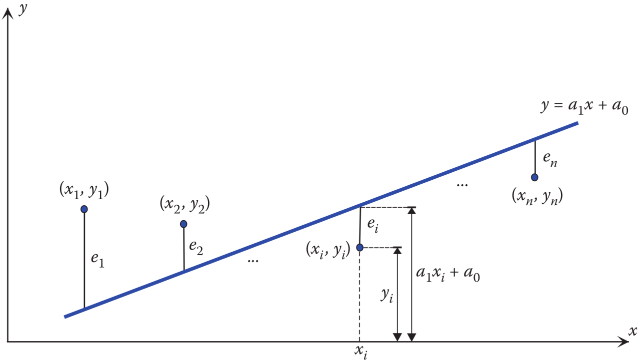
- On a small data set, a single polynomial may not be sufficient.
- When the data has a substantial error.



Least Squares Approximation

- least-squares approximation involves finding a straight line in the form
$$y = a_1 x + a_0$$
- that best fits a set of n data points (x_i, y_i), for $i = 1, 2, \cdots n$.
- at each data point (x_i, y_i) the error e_i is defined as

$$e_i = y_i - (a_1 x_i + a_0)$$



How to decide a BEST FIT

- Minimize the sum of all the individual errors

$$E \equiv \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} [y_i - (a_1 x_i + a_0)]$$

How to decide a BEST FIT

- Minimize the sum of all the individual errors

$$E \equiv \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} [y_i - (a_1 x_i + a_0)]$$

Zero-sum: Positive and negative individual errors-even very large errors-to cancel out.

- Minimize the sum of the absolute values of the individual errors

$$E \equiv \sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - (a_1 x_i + a_0)|$$

How to decide a BEST FIT

- Minimize the sum of all the individual errors

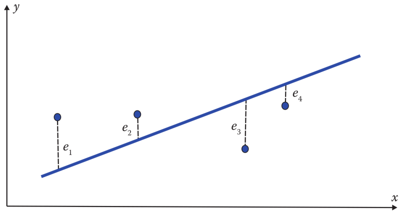
$$E \equiv \sum_{i=1}^{n} e_i = \sum_{i=1}^{n} [y_i - (a_1 x_i + a_0)]$$

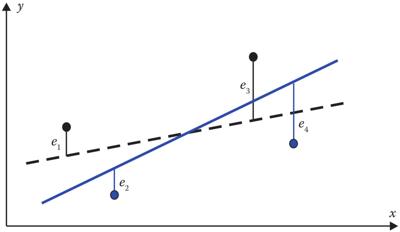
Zero-sum: Positive and negative individual errors-even very large errors-to cancel out.

- Minimize the sum of the absolute values of the individual errors

$$E \equiv \sum_{i=1}^{n} |e_i| = \sum_{i=1}^{n} |y_i - (a_1 x_i + a_0)|$$

The individual errors can no longer cancel out, and the total error is always positive.





Linear Least Squares

- Let there be a collection of data $\{(x_i, y_i)\}_{i=1}^{m}$
- To fit the best least squares line to the given collection of data requires minimizing sum of the squares of the individual errors.

$$E \equiv E(a_0, a_1) = \sum_{i=1}^{m} e_i^2 = \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2$$

where the parameters a_0 and a_1 are such that

$$\frac{\partial E}{\partial a_0} = \frac{\partial}{\partial a_0} \sum_{i=1}^{m} [(y_i - (a_1 x_i + a_0)]^2 = -2 \sum_{i=1}^{m} (y_i - a_1 x_i - a_0) = 0,$$

$$\frac{\partial E}{\partial a_1} = \frac{\partial}{\partial a_1} \sum_{i=1}^{m} [y_i - (a_1 x_i + a_0)]^2 = -2 \sum_{i=1}^{m} (y_i - a_1 x_i - a_0) x_i = 0.$$

Linear Least Squares

- Simplifying the above two equations gives us the normal equations

$$
\begin{aligned}
a_0 \cdot m + a_1 \sum_{i=1}^{m} x_i &= \sum_{i=1}^{m} y_i, \\
a_0 \sum_{i=1}^{m} x_i + a_1 \sum_{i=1}^{m} x_i^2 &= \sum_{i=1}^{m} x_i y_i.
\end{aligned}
$$

- Solving the two equations, we get

$$
\begin{aligned}
a_0 &= \frac{\sum_{i=1}^{m} x_i^2 \sum_{i=1}^{m} y_i - \sum_{i=1}^{m} x_i y_i \sum_{i=1}^{m} x_i}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}, \quad \text{and} \\
a_1 &= \frac{m \sum_{i=1}^{m} x_i y_i - \sum_{i=1}^{m} x_i \sum_{i=1}^{m} y_i}{m\left(\sum_{i=1}^{m} x_i^2\right) - \left(\sum_{i=1}^{m} x_i\right)^2}.
\end{aligned}
$$

Example

Example

Find the least square straight line fit to the following data

x	0	2	5	7
$f(x)$	1	5	12	20

Example

Example

Find the least square straight line fit to the following data

x	0	2	5	7
$f(x)$	1	5	12	20

Ans: $P_1(x) = -1.1381 + 2.8966x$

$$\text{Least squares error} = \sum_{i=0}^{4}[f(x_i) - (-1.1381 + 2.8966x_i)]^2 = ?$$


ANY
QUESTIONS?