Newton's method

The theory (a short excerpts from lectures)

The goal: find a solution of a system of nonlinear equations $\mathbf{F}(\mathbf{x}) = \mathbf{0}$, where

$$\mathbf{F}(\mathbf{x}) : R^n \to R^n \ , \quad \mathbf{x} = \begin{bmatrix} x_1 \\ \dots \\ x_n \end{bmatrix}, \quad \mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(x_1, ...x_n) \\ \dots \\ f_n(x_1, ...x_n) \end{bmatrix}, \quad f_i : R^n \to R \text{ differentiable.}$$

The idea: for a given approximation $\widehat{\mathbf{x}}$ of the solution, linearize the equations around $\widehat{\mathbf{x}}$ and find a better approximation as the exact solution of these linearized equations. Repeat the process until convergence.

Derivation of the algorithm for $n = 2$ (the generalization is straightforward):

$$f_1(\mathbf{x}) \approx f_1(\widehat{\mathbf{x}}) + \frac{\partial f_1}{\partial x_1}(\widehat{\mathbf{x}}) \cdot (x_1 - \widehat{x}_1) + \frac{\partial f_1}{\partial x_2}(\widehat{\mathbf{x}}) \cdot (x_2 - \widehat{x}_2)$$

$$f_2(\mathbf{x}) \approx f_2(\widehat{\mathbf{x}}) + \frac{\partial f_2}{\partial x_1}(\widehat{\mathbf{x}}) \cdot (x_1 - \widehat{x}_1) + \frac{\partial f_2}{\partial x_2}(\widehat{\mathbf{x}}) \cdot (x_2 - \widehat{x}_2)$$

Instead of the nonlinear system $f_1(\mathbf{x}) = 0$ and $f_2(\mathbf{x}) = 0$, solve the linear one:

$$f_1(\widehat{\mathbf{x}}) + \frac{\partial f_1}{\partial x_1}(\widehat{\mathbf{x}}) \cdot (x_1 - \widehat{x}_1) + \frac{\partial f_1}{\partial x_2}(\widehat{\mathbf{x}}) \cdot (x_2 - \widehat{x}_2) = 0$$

$$f_2(\widehat{\mathbf{x}}) + \frac{\partial f_2}{\partial x_1}(\widehat{\mathbf{x}}) \cdot (x_1 - \widehat{x}_1) + \frac{\partial f_2}{\partial x_2}(\widehat{\mathbf{x}}) \cdot (x_2 - \widehat{x}_2) = 0$$

which in matrix form can be rewritten as

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\widehat{\mathbf{x}}) & \frac{\partial f_1}{\partial x_2}(\widehat{\mathbf{x}}) \\ \frac{\partial f_2}{\partial x_1}(\widehat{\mathbf{x}}) & \frac{\partial f_2}{\partial x_2}(\widehat{\mathbf{x}}) \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} = - \begin{bmatrix} f_1(\widehat{\mathbf{x}}) \\ f_2(\widehat{\mathbf{x}}) \end{bmatrix} , \quad \text{where } \mathbf{d} = \mathbf{x} - \widehat{\mathbf{x}}$$

Algorithm of Newton's method

1. compute Jacobi matrix

$$\mathbf{F}'(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

© Certik

2. choose $\mathbf{x}^{(0)}$
3. repeat for $k = 0, 1, 2, \ldots$

   3.1 compute a vector $\mathbf{d}^{(k)}$ as the solution of the system

   $$\mathbf{F}'(\mathbf{x}^{(k)})\,\mathbf{d}^{(k)} = -\mathbf{F}(\mathbf{x}^{(k)})$$

   3.2 set $\quad \mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \mathbf{d}^{(k)}$

   until $\|\mathbf{F}(\mathbf{x}^{(k+1)})\| < \varepsilon$ and $\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| < \varepsilon$,
   or $k > k_{max}$ for some ε and k_{max} of your choice.

For all vectors $\mathbf{x}^{(k)}$, the matrix $\mathbf{F}'(\mathbf{x}^{(k)})$ has to be nonsingular, so that the system 3.1 has a unique solution. If for some k the matrix $\mathbf{F}'(\mathbf{x}^{(k)})$ happens to be singular, choose different vector $\mathbf{x}^{(0)}$ and start the process again.

Theorem - convergence of Newton's method:

Assume that

- $\mathbf{F}$ is continuously differentiable twice in a domain $D \subset R^n$,
- there exists a solution $\mathbf{x}^* \in D$ of the system $\mathbf{F}(\mathbf{x}) = \mathbf{0}$,
- $\mathbf{F}'(\mathbf{x}^*)$ is nonsingular (i.e., invertible).

Then there exists a neighbourhood U_δ of $\mathbf{x}^*$ such that Newton's method covnerges for any starting point $\mathbf{x}^{(0)} \in U_\delta$. Moreover, the convergence is quadratic: there exists a constant c such that

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^*\| \le c\,\|\mathbf{x}^{(k)} - \mathbf{x}^*\|^2.$$

However, in practice usually we are not able to localize the solution $\mathbf{x}^*$ with precision δ, we even do not know what δ is (i.e., how close to the solution the starting point $\mathbf{x}^{(0)}$ should be), and we do not know if $\mathbf{F}'(\mathbf{x}^*)$ is invertible, either.
So, in practice, Newton's method is typically used in a "trial – error" way.

Example 1
Consider the following system of nonlinear equations:

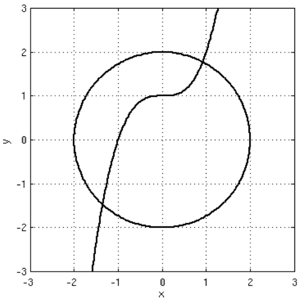
$$\begin{aligned} x_1^2 + x_2^2 &= 4 \\ x_2 &= x_1^3 + 1 \end{aligned}$$

a) Find the solution graphically.
b) Choose $\mathbf{x}^{(0)} = (\,1,\,2)^T$ and compute the first two iterations of Newton's method.
c) Can $\mathbf{x}^{(0)}$ be choosen as $\mathbf{x}^{(0)} = (\,1,\,-1/3)^T$? Give reasons for your answer.

© Certik

The solution

a)



b) Transform the system to the vector notation

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} f_1(x_1,x_2) \\ \\ f_2(x_1,x_2) \end{bmatrix} = \begin{bmatrix} x_1^2 + x_2^2 - 4 \\ \\ x_2 - x_1^3 - 1 \end{bmatrix} = \begin{bmatrix} 0 \\ \\ 0 \end{bmatrix} \equiv \mathbf{0}\,, \quad \text{where } \mathbf{x} = \begin{bmatrix} x_1 \\ \\ x_2 \end{bmatrix}$$

and compute the Jacobian matrix

$$\mathbf{F}'(\mathbf{x}) = \begin{bmatrix} \dfrac{\partial f_1}{\partial x_1} & \dfrac{\partial f_1}{\partial x_2} \\ \\ \dfrac{\partial f_2}{\partial x_1} & \dfrac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 2x_1 & 2x_2 \\ \\ -3x_1^2 & 1 \end{bmatrix}$$

The first iteration of Newton's method:

- compute the vector $\mathbf{d}^{(0)} = (d_1^{(0)}, d_2^{(0)})^T$ by solving the system $\mathbf{F}'(\mathbf{x}^{(0)})\,\mathbf{d}^{(0)} = -\mathbf{F}(\mathbf{x}^{(0)})$:

  substitution of $\mathbf{x}^{(0)} = (\,1,\,2)^T$ leads to

$$\mathbf{F}'(\mathbf{x}^{(0)}) = \begin{bmatrix} 2\cdot 1 & 2\cdot 2 \\ \\ -3\cdot 1^2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ \\ -3 & 1 \end{bmatrix}$$

$$\mathbf{F}(\mathbf{x}^{(0)}) = \begin{bmatrix} 1^2 + 2^2 - 4 \\ \\ 2 - 1^3 - 1 \end{bmatrix} = \begin{bmatrix} 1 \\ \\ 0 \end{bmatrix}$$

  now $\mathbf{d}^{(0)}$ is to be computed from the linear system

$$\begin{bmatrix} 2 & 4 \\ \\ -3 & 1 \end{bmatrix} \begin{bmatrix} d_1^{(0)} \\ \\ d_2^{(0)} \end{bmatrix} = -\begin{bmatrix} 1 \\ \\ 0 \end{bmatrix} \quad \Rightarrow \quad \mathbf{d}^{(0)} = \begin{bmatrix} -1/14 \\ \\ -3/14 \end{bmatrix}$$

- compute $\mathbf{x}^{(1)}$:

$$\mathbf{x}^{(1)} = \mathbf{x}^{(0)} + \mathbf{d}^{(0)} = \begin{bmatrix} 1 \\ \\ 2 \end{bmatrix} + \begin{bmatrix} -1/14 \\ \\ -3/14 \end{bmatrix} = \begin{bmatrix} 13/14 \\ \\ 25/14 \end{bmatrix} = \begin{bmatrix} 0.9286 \\ \\ 1.7857 \end{bmatrix}$$

© Certik

The second iteration of Newton's method:

- compute the vector $\mathbf{d}^{(1)} = (d_1^{(1)}, d_2^{(1)})^T$ by solving the system $\mathbf{F}'(\mathbf{x}^{(1)})\,\mathbf{d}^{(1)} = -\mathbf{F}(\mathbf{x}^{(1)})$:

  substitution $\mathbf{x}^{(1)} = (\,13/14,\ 25/14)^T$ leads to

$$\mathbf{F}'(\mathbf{x}^{(1)}) = \begin{bmatrix} 2\cdot 13/14 & 2\cdot 25/14 \\ -3\cdot(13/14)^2 & 1 \end{bmatrix} = \begin{bmatrix} 13/7 & 25/7 \\ -507/196 & 1 \end{bmatrix}$$

$$\mathbf{F}(\mathbf{x}^{(1)}) = \begin{bmatrix} (13/14)^2 + (25/14)^2 - 4 \\ 25/14 - (13/14)^3 - 1 \end{bmatrix} = \begin{bmatrix} 5/98 \\ -41/2744 \end{bmatrix}$$

now $\mathbf{d}^{(1)}$ is to be computed from the linear system

$$\begin{bmatrix} 13/7 & 25/7 \\ -507/196 & 1 \end{bmatrix} \begin{bmatrix} d_1^{(1)} \\ d_2^{(1)} \end{bmatrix} = - \begin{bmatrix} 5/98 \\ -41/2744 \end{bmatrix}$$

$$\Rightarrow \quad \mathbf{d}^{(1)} = \begin{bmatrix} -105/11161 \\ -11/1171 \end{bmatrix} = \begin{bmatrix} -0.009408 \\ -0.009394 \end{bmatrix}$$

- compute $\mathbf{x}^{(2)}$:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} + \mathbf{d}^{(1)} = \begin{bmatrix} 13/14 \\ 25/14 \end{bmatrix} + \begin{bmatrix} -105/11161 \\ -11/1171 \end{bmatrix} = \begin{bmatrix} 1319/1435 \\ 4876/2745 \end{bmatrix} = \begin{bmatrix} 0.9192 \\ 1.7763 \end{bmatrix}$$

c) Compute $\mathbf{F}'(\mathbf{x}^{(0)})$:

$$\mathbf{F}'(\mathbf{x}^{(0)}) = \begin{bmatrix} 2\cdot 1 & 2\cdot(-1/3) \\ -3\cdot 1^2 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -2/3 \\ -3 & 1 \end{bmatrix}$$

This matrix is singular (the rows are linearly dependent), so the equation at the item 3.1 of Newton's method does not have a unique solution. In this case the initial approximation $\mathbf{x}^{(0)}$ has to be chosen differently.

© Certik