Chapter 8 Numerical Solutions of Nonlinear Systems of Equations

Numerical Solutions of Nonlinear Systems of Equations

Newton method:

The iteration mapping function, $\mathbf{G}(\mathbf{x}) = \mathbf{x} - J(\mathbf{x})^{-1}\mathbf{F}(\mathbf{x})$, where

$$J(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1(\mathbf{x})}{\partial x_1} & \frac{\partial f_1(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_1(\mathbf{x})}{\partial x_n} \\ \frac{\partial f_2(\mathbf{x})}{\partial x_1} & \frac{\partial f_2(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_2(\mathbf{x})}{\partial x_n} \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n(\mathbf{x})}{\partial x_1} & \frac{\partial f_n(\mathbf{x})}{\partial x_2} & \cdots & \frac{\partial f_n(\mathbf{x})}{\partial x_n} \end{bmatrix}.$$

The solution of $\mathbf{x}$ can be approximated by the iteration

$\mathbf{x}^{(k)} = \mathbf{G}(\mathbf{x}^{(k-1)}) = \mathbf{x}^{(k-1)} - J(\mathbf{x}^{(k-1)})^{-1}\mathbf{F}(\mathbf{x}^{(k-1)})$.

Quasi-Newton method:

The secant method uses the approximation

$$f'(x_1) \approx \frac{f(x_1) - f(x_0)}{x_1 - x_0}$$

as a replacement for $f'(x_1)$ in Newton's method. For nonlinear systems, $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ is a vector, and the corresponding quotient is undefined. However, the method proceeds similarly in that we replace the matrix $J(\mathbf{x}^{(1)})$ in Newton's method by a matrix A_1 with the property that

$$A_1(\mathbf{x}^{(1)} - \mathbf{x}^{(0)}) = \mathbf{F}(\mathbf{x}^{(1)}) - \mathbf{F}(\mathbf{x}^{(0)}) . \qquad (1)$$

This equation does not define a unique matrix, because it does not describe how A_1 operates on vectors orthogonal to $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$. Since no information is available about the change in $\mathbf{F}$ in a direction orthogonal to $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$, we require additionally of A_1 that

$$A_1\mathbf{z} = J(\mathbf{x}^{(0)})\mathbf{z}, \qquad \text{whenever } (\mathbf{x}^{(1)} - \mathbf{x}^{(0)})^t \mathbf{z} = 0 . \qquad (2)$$

This condition specifies that any vector orthogonal to $\mathbf{x}^{(1)} - \mathbf{x}^{(0)}$ is unaffected by the update from $J(\mathbf{x}^{(0)})$, which was used to compute $\mathbf{x}^{(1)}$, to A_1 ,which is used in the determination of $\mathbf{x}^{(2)}$.

Conditions (1) and (2) uniquely define A_1, as

$$A_1 = J(\mathbf{x}^{(0)}) + \frac{[\mathbf{F}(\mathbf{x}^{(1)}) - \mathbf{F}(\mathbf{x}^{(0)}) - J(\mathbf{x}^{(0)})(\mathbf{x}^{(1)} - \mathbf{x}^{(0)})](\mathbf{x}^{(1)} - \mathbf{x}^{(0)})^t}{\left\|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\right\|_2^2} .$$

It is the matrix that is used in place of $J(\mathbf{x}^{(1)})$ to determine $\mathbf{x}^{(2)}$:

$$\mathbf{x}^{(2)} = \mathbf{x}^{(1)} - A_1^{-1}\mathbf{F}(\mathbf{x}^{(1)}) .$$

This method is then repeated to determine $\mathbf{x}^{(3)}$, using A_1 in place of $A_0 \equiv J(\mathbf{x}^{(0)})$ and with $\mathbf{x}^{(2)}$ and $\mathbf{x}^{(1)}$ in place of $\mathbf{x}^{(1)}$ and $\mathbf{x}^{(0)}$. In general, once $\mathbf{x}^{(i)}$ has been determined, $\mathbf{x}^{(i+1)}$ is computed by

$$A_i = A_{i-1} + \frac{\mathbf{y}_i - A_{i-1}\mathbf{s}_i}{\left\|\mathbf{s}_i\right\|_2^2}\mathbf{s}_i^t ,$$

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - A_i^{-1}\mathbf{F}(\mathbf{x}^{(i)}) ,$$

where the notation $\mathbf{y}_i = \mathbf{F}(\mathbf{x}^{(i)}) - \mathbf{F}(\mathbf{x}^{(i-1)})$ and $\mathbf{s}_i = \mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}$.

Steepest Descent:

The method of Steepest Descent determines a local minimum for a multivariable function of the form $g:\mathbf{R}^n \to \mathbf{R}$. The connection between the minimization of a function from $\mathbf{R}^n$ to $\mathbf{R}$ and the solution of a system of nonlinear equations is due to the fact that a system of the form

$$f_1(x_1,x_2,\ldots,x_n)=0,$$
$$f_2(x_1,x_2,\ldots,x_n)=0,$$
$$\vdots$$
$$f_n(x_1,x_2,\ldots,x_n)=0,$$

has a solution at $\mathbf{x}=(x_1,x_2,\ldots,x_n)^t$ precisely when the function g defined by

$$g(x_1,x_2,\ldots,x_n)=\sum_{i=1}^{n}[f_i(x_1,x_2,\ldots,x_n]^2$$

has the minimal value zero.

The method of Steepest Descent for finding a local minimum for an arbitrary function g from $\mathbf{R}^n$ into $\mathbf{R}$ can be intuitively described as follows:

i. Evaluate g at an initial approximation $\mathbf{x}^{(0)}=(x_1^{(0)},x_2^{(0)},\ldots,x_n^{(0)})^t$;

ii. Determine a direction from $\mathbf{x}^{(0)}$ that results in a decrease in the value of g;

iii. Move an appropriate distance in this direction and call the new vector $\mathbf{x}^{(1)}$;

iv. Repeat steps i through iii with $\mathbf{x}^{(0)}$ by $\mathbf{x}^{(1)}$.

An appropriate choice for $\mathbf{x}^{(1)}$ is

$$\mathbf{x}^{(1)}=\mathbf{x}^{(0)}-\alpha\nabla g(\mathbf{x}^{(0)}), \qquad \text{for some constant } \alpha>0.$$

To determine an appropriate choice of α, we consider the single-variable function

$$h(\alpha)=g(\mathbf{x}^{(0)}-\alpha\nabla g(\mathbf{x}^{(0)}).$$

The value of α that minimize $h(\alpha)$ is what we need.

References:

【1】 R. L. Burden and J. D. Faires, *Numerical Analysis*, PWS, Boston, 1993.