

Solution of systems of nonlinear equations

Given a system of nonlinear equations

$$\mathbf{f}(\mathbf{x}) = 0, \quad \mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^m \quad (1)$$

for which we assume that there is (at least) one solution \mathbf{x}^* . The idea is to rewrite this system into the form

$$\mathbf{x} = \mathbf{g}(\mathbf{x}), \quad \mathbf{g} : \mathbb{R}^m \rightarrow \mathbb{R}^m. \quad (2)$$

The solution ξ of (1) should satisfy $\xi = \mathbf{g}(\xi)$, and is thus called a *fixed point* of \mathbf{g} . The iteration schemes becomes: given an initial guess $\mathbf{x}^{(0)}$, the *fixed point iterations* becomes

$$\mathbf{x}^{(k+1)} = \mathbf{g}(\mathbf{x}^{(k)}), \quad k = 1, 2, \dots \quad (3)$$

The following questions arise:

- (i) How to find a suitable function \mathbf{g} ?
- (ii) Under what conditions will the sequence $\mathbf{x}^{(k)}$ converge to the fixed point ξ ?
- (iii) How quickly will the sequence $\mathbf{x}^{(k)}$ converge?

Point (ii) can be answered by Banach fixed point theorem:

Theorem 0.1. *Let $D \subseteq \mathbb{R}^m$ be a convex¹ and closed set. If*

$$\mathbf{g}(D) \subseteq D \quad (4a)$$

and

$$\|\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{v})\| \leq L\|\mathbf{y} - \mathbf{v}\|, \quad \text{with } L < 1 \text{ for all } \mathbf{y}, \mathbf{v} \in D, \quad (4b)$$

then G has a unique fixed point in D and the fixed point iterations (3) converges for all $\mathbf{x}^{(0)} \in D$. Further,

$$\|\mathbf{x}^{(k)} - \xi\| \leq \frac{L^k}{1-L} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|. \quad (4c)$$

Proof. The proof is based on the *Cauchy Convergence theorem*, saying that a sequence $\{\mathbf{x}^{(k)}\}_{k=0}^{\infty}$ converges to some ξ if and only if for every $\varepsilon > 0$ there is an N such that

$$\|\mathbf{x}^{(l)} - \mathbf{x}^{(k)}\| < \varepsilon \quad \text{for all } l, k > N. \quad (5)$$

Assumption (4a) ensures $\mathbf{x}^{(k)} \in D$ as long as $\mathbf{x}^{(0)} \in D$. From (3) and (4b) we get:

$$\|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| = \|\mathbf{g}(\mathbf{x}^{(k)}) - \mathbf{g}(\mathbf{x}^{(k-1)})\| \leq L\|\mathbf{x}^{(k)} - \mathbf{x}^{(k-1)}\| \leq L^k \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|.$$

We can write $\mathbf{x}^{(k+p)} - \mathbf{x}^{(k)} = \sum_{i=1}^p (\mathbf{x}^{(k+i)} - \mathbf{x}^{(k+i-1)})$, thus

$$\begin{aligned} \|\mathbf{x}^{(k+p)} - \mathbf{x}^{(k)}\| &\leq \sum_{i=1}^p \|\mathbf{x}^{(k+i)} - \mathbf{x}^{(k+i-1)}\| \\ &= (L^{p-1} + L^{p-2} + \dots + 1) \|\mathbf{x}^{(k+1)} - \mathbf{x}^{(k)}\| \leq \frac{L^k}{1-L} \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|, \end{aligned}$$

¹ D is convex if $\theta y + (1 - \theta)v \in D$ for all $y, v \in D$ and $\theta \in [0, 1]$.

since $L < 1$. For the same reason, the sequence satisfy (5), so the sequence converges to some $\xi \in D$. Since the inequality is true for all $p > 0$ it is also true for ξ , proving (4c).

To prove that the fixed point is unique, let ξ and η be two different fixed points in D . Then

$$\|\xi - \eta\| = \|\mathbf{g}(\xi) - \mathbf{g}(\eta)\| < \|\xi - \eta\|$$

which is impossible. □

For a given problem, it is not necessarily straightforward to justify the two assumptions of the theorem. But it is sufficient to find some L satisfying the condition $L < 1$ in some norm to prove convergence.

Let $\mathbf{x} = [x_1, \dots, x_m]^T$ and $\mathbf{g}(\mathbf{x}) = [g_1(\mathbf{x}), \dots, g_m(\mathbf{x})]^T$. Let $\mathbf{y}, \mathbf{v} \in D$, and let $\mathbf{x}(\theta) = \theta\mathbf{y} + (1 - \theta)\mathbf{v}$ be the straight line between \mathbf{y} and \mathbf{v} . According to the mean value theorem for functions, for each g_i there exist at $\tilde{\theta}_i$ such that

$$\begin{aligned} g_i(\mathbf{y}) - g_i(\mathbf{v}) &= g_i(\mathbf{x}(1)) - g_i(\mathbf{x}(0)) = \frac{dg_i}{d\theta}(\tilde{\theta}_i)(1 - 0), & \tilde{\theta}_i &\in (0, 1) \\ &= \sum_{j=1}^m \frac{\partial g_i}{\partial x_j}(\tilde{\mathbf{x}}_i)(y_j - v_j), & \tilde{\mathbf{x}}_i &= \tilde{\theta}_i\mathbf{y} + (1 - \tilde{\theta}_i)\mathbf{v} \end{aligned}$$

since $dx_j(\theta)/d\theta = y_j - v_j$. Then

$$|g_i(\mathbf{y}) - g_i(\mathbf{v})| \leq \sum_{j=1}^m \left| \frac{\partial g_i}{\partial x_j}(\tilde{\mathbf{x}}_i) \right| \cdot |y_j - v_j| \leq \left(\sum_{j=1}^m \left| \frac{\partial g_i}{\partial x_j}(\tilde{\mathbf{x}}_i) \right| \right) \max_l |y_l - v_l|.$$

If we let \bar{g}_{ij} be some upper bound for each of the partial derivatives, that is

$$\left| \frac{\partial g_i}{\partial x_j}(\mathbf{x}) \right| \leq \bar{g}_{ij}, \quad \text{for all } \mathbf{x} \in D.$$

then

$$\|\mathbf{g}(\mathbf{y}) - \mathbf{g}(\mathbf{v})\|_\infty = \left(\max_i \sum_{j=1}^m \bar{g}_{ij} \right) \|\mathbf{y} - \mathbf{v}\|_\infty.$$

We can then conclude that (4b) is satisfied if

$$\max_i \sum_{j=1}^m \bar{g}_{ij} < 1.$$

Newton's method

Newton's method is a fixed point iterations for which

$$\mathbf{g}(\mathbf{x}^{(k)}) = \mathbf{x}^{(k)} - J_f(\mathbf{x}^{(k)})^{-1} \mathbf{f}(\mathbf{x}^{(k)}), \quad (6)$$

where the *Jacobian* is the matrix function

$$J_f(\mathbf{x}) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_1}{\partial x_m}(\mathbf{x}) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(\mathbf{x}) & \cdots & \frac{\partial f_m}{\partial x_m}(\mathbf{x}) \end{pmatrix}.$$

The Newton method can be derived as follow: Consider element i in \mathbf{f} , that is $f_i(\mathbf{x})$. Do a multidimensional Taylor expansion of $f_i(\xi)$ around the vector $\mathbf{x}^{(k)}$, using $\mathbf{e}^{(k)} = \xi - \mathbf{x}^{(k)}$. This gives

$$0 = f_i(x_1^{(k)} + e_1^{(k)}, \dots, x_m^{(k)} + e_m^{(k)}) = f_i + \frac{\partial f_i}{\partial x_1} e_1^{(k)} + \dots + \frac{\partial f_i}{\partial x_m} e_m^{(k)} + R_i$$

The function and all the derivatives are evaluated in $\mathbf{x}^{(k)}$. The remainder term R_i consists of quadratic terms like $\mathcal{O}(e_i^{(k)} e_j^{(k)})$. If the error is small, this term is even smaller, so let us now ignore it and replace the errors $e_i^{(k)}$ with an approximation to the error $\Delta x_i^{(k)}$ to compensate. Doing so for each $i = 1, 2, \dots, m$ gives us the following system of linear equations,

$$f_i + \frac{\partial f_i}{\partial x_1} \Delta x_1^{(k)} + \dots + \frac{\partial f_i}{\partial x_m} \Delta x_m^{(k)} = 0, \quad i = 1, 2, \dots, m.$$

which is

$$\mathbf{f}(\mathbf{x}^{(k)}) + J_f(\mathbf{x}^{(k)}) \cdot \Delta \mathbf{x}^{(k)} = \mathbf{0}.$$

Solve this with respect to $\Delta \mathbf{x}^{(k)}$. Remember that $\Delta \mathbf{x}^{(k)} \approx \xi - \mathbf{x}_k^{(k)}$ it seems reasonable to update our iterate with this amount, thus

$$\mathbf{x}^{(k+1)} = \mathbf{x}^{(k)} + \Delta \mathbf{x}_k^{(k)}$$

which finally results in (6).

It is possible to prove that if $i)$ (1) has a solution ξ , $ii)$ $J_f(\mathbf{x})$ is nonsingular in some open neighbourhood around ξ and $iii)$ the initial guess $\mathbf{x}^{(0)}$ is sufficiently close to ξ , the Newton iterations will converge to ξ and

$$\|\xi - \mathbf{x}^{(k+1)}\| \leq K \|\xi - \mathbf{x}^{(k)}\|^2$$

for some positive constant K . We say that the convergence is *quadratic*.