**Theorem 4.2** *Suppose that* $\boldsymbol{g} = (g_1, \ldots, g_n)^{\mathrm{T}} \colon \mathbb{R}^n \to \mathbb{R}^n$ *is defined and continuous on a closed set* $D \subset \mathbb{R}^n$. *Let* $\boldsymbol{\xi} \in D$ *be a fixed point of* $\boldsymbol{g}$, *and suppose that the first partial derivatives* $\frac{\partial g_i}{\partial x_j}$, $j = 1, \ldots, n$, *of* $g_i$, $i = 1, \ldots, n$, *are defined and continuous in some (open) neighbourhood* $N(\boldsymbol{\xi}) \subset D$ *of* $\boldsymbol{\xi}$, *with*

$$\|Dg(\boldsymbol{\xi})\|_\infty < 1 .$$

*Then, there exists* $\varepsilon > 0$ *such that* $\boldsymbol{g}(\bar{B}_\varepsilon(\boldsymbol{\xi})) \subset \bar{B}_\varepsilon(\boldsymbol{\xi})$, *and the sequence defined by (4.3) converges to* $\boldsymbol{\xi}$ *for all* $\boldsymbol{x}^{(0)} \in \bar{B}_\varepsilon(\boldsymbol{\xi})$.

Newton's Method

$$g(x) = x - (Df(x))^{-1} f(x)$$

$$Dg(x) = Dx - D\left((Df(x))^{-1} f(x)\right) = I - D\left((Df(x))^{-1} f(x)\right)$$

$$\left[D\left((Df(x))^{-1} f(x)\right)\right]_{ij} = \frac{\partial}{\partial x_j} \sum_{r=1}^{n} K_{ir}(x) f_r(x)$$

where $K(x) = [Df(x)]^{-1}$

need to verify this is cont. func. of $x$.

$$\ldots = \sum_{r=1}^{n} \frac{\partial K_{ir}(x)}{\partial x_j} f_r(x) + \sum_{r=1}^{n} K_{ir}(x) \frac{\partial f_r(x)}{\partial x_j}$$

$$\frac{\partial}{\partial x_j}\left[D\left((Df(x))^{-1} f(x)\right)\right]_i \Bigg|_{x=\xi} = \sum_{r=1}^{n} K_{ir}(\xi) \frac{\partial f_r(\xi)}{\partial x_j} = \left[[Df(\xi)]^{-1} Df(\xi)\right]_{ij}$$

$$= I_{ij}$$

In other words

$$D\left((Df(\xi))^{-1}f(\xi)\right) = I$$

$$Dg(\xi) = I - D\left((Df(\xi))^{-1}f(\xi)\right) = I - I = 0$$

the zero matrix

$$\|Dg(\xi)\| = 0.$$

Since $Dg(x)$ is continuous, ... why is it cont.

**Theorem 4.4** *Suppose that* $\boldsymbol{f}(\boldsymbol{\xi}) = \mathbf{0}$, *that in some (open) neighbourhood* $N(\boldsymbol{\xi})$ *of* $\boldsymbol{\xi}$, *where* $\boldsymbol{f}$ *is defined and continuous, all the second-order partial derivatives of* $\boldsymbol{f}$ *are defined and continuous, and that the Jacobian matrix* $J_{\boldsymbol{f}}(\boldsymbol{\xi})$ *of* $\boldsymbol{f}$ *at the point* $\boldsymbol{\xi}$ *is nonsingular. Then, the sequence* $(\boldsymbol{x}^{(k)})$ *defined by Newton's method (4.18) converges to the solution* $\boldsymbol{\xi}$ *provided that* $\boldsymbol{x}^{(0)}$ *is sufficiently close to* $\boldsymbol{\xi}$*; the convergence of the sequence* $(\boldsymbol{x}^{(k)})$ *to* $\boldsymbol{\xi}$ *is at least quadratic.*

$$\left[D\left((Df(x))^{-1}f(x)\right)\right]_{ij} = \sum_{r=1}^{m} \frac{\partial K_{ir}(x)}{\partial x_j} f_r(x) + \sum_{r=1}^{m} K_{ir}(x) \frac{\partial f_r(x)}{\partial x_j}$$

consists of second order partial derivative

this is continuous

if the second order partial derivatives are defined then second order must be continuous

$$K(x) = [Df(x)]^{-1}$$

So $K$ is made out of partial derivatives of $f$.

Hypothesis satisfied implies there exists

*Then, there exists* $\varepsilon > 0$ *such that* $\boldsymbol{g}(\bar{B}_\varepsilon(\boldsymbol{\xi})) \subset \bar{B}_\varepsilon(\boldsymbol{\xi})$, *and the sequence defined by (4.3) converges to* $\boldsymbol{\xi}$ *for all* $\boldsymbol{x}^{(0)} \in \bar{B}_\varepsilon(\boldsymbol{\xi})$.

What is the rate of convergence...

**Theorem 4.4** *Suppose that* $\boldsymbol{f}(\boldsymbol{\xi}) = \mathbf{0}$, *that in some (open) neighbourhood* $N(\boldsymbol{\xi})$ *of* $\boldsymbol{\xi}$, *where* $\boldsymbol{f}$ *is defined and continuous, all the second-order partial derivatives of* $\boldsymbol{f}$ *are defined and continuous, and that the Jacobian matrix* $D\boldsymbol{f}(\boldsymbol{\xi})$ *of* $\boldsymbol{f}$ *at the point* $\boldsymbol{\xi}$ *is nonsingular. Then, the sequence* $(\boldsymbol{x}^{(k)})$ *defined by Newton's method (4.18) converges to the solution* $\boldsymbol{\xi}$ *provided that* $\boldsymbol{x}^{(0)}$ *is sufficiently close to* $\boldsymbol{\xi}$; *the convergence of the sequence* $(\boldsymbol{x}^{(k)})$ *to* $\boldsymbol{\xi}$ *is at least quadratic.*

Let $e_k = x_k - \xi$.

$$e_{k+1} = x_{k+1} - \xi = g(x_k) - \xi = x_k - Df(x_k)^{-1} f(x_k) - \xi$$

$$= e_k - Df(x_k)^{-1} f(x_k)$$

Taylor's theorem in $\mathbb{R}^n$

since $\xi$ is a root

$$0 = f(\xi) = f(x) + Df(x)(\xi - x) + E_1$$

Theorem says

$$|E_1| \leq \frac{1}{2} A n^2 \|\xi - x\|_\infty^2$$

**Theorem A.7 (Taylor's Theorem for several variables)** *Suppose that $f$ is a real-valued function of $n$ real variables, $n \geq 1$, such that $f$ and all of its partial derivatives up to and including order $k+1$ are defined, continuous and bounded in a neighbourhood of the point $\boldsymbol{a}$ in $\mathbb{R}^n$. Let $A$ denote an upper bound on the absolute values of all the derivatives of order $k+1$ in this neighbourhood. Then*

$$f(\boldsymbol{a}+\boldsymbol{\eta}) = f(\boldsymbol{a}) + \sum_{r=1}^{k} \frac{U_r(\boldsymbol{a})}{r!} + E_k\,,$$

*where*

$$U_r(\boldsymbol{a}) = \left[\left(\eta_1 \frac{\partial}{\partial x_1} + \cdots + \eta_n \frac{\partial}{\partial x_n}\right)^r f\right](\boldsymbol{a})\,, \qquad r = 1, \ldots, k\,,$$

*and*

$$|E_k| \leq \tfrac{1}{(k+1)!} A n^{k+1} \|\boldsymbol{\eta}\|_\infty^{k+1}.$$

since $\xi$ is a root

$$0 = f(\xi) = f(x^{(k)}) + Df(x^{(k)})(\underbrace{\xi - x^{(k)}}_{-e_k}) + E_1$$

$$0 = f(x^{(k)}) - \underbrace{Df(x^{(k)})}_{\text{this is invertible since } x^{(k)} \in \overline{B}_\varepsilon(\xi)} e_k + E_1$$

$$0 = Df(x^{(k)})^{-1} f(x^{(k)}) - e_k + Df(x^{(k)})^{-1} E_1$$

$$-Df(x^{(k)})^{-1} f(x^{(k)}) = -e_k + Df(x^{(k)})^{-1} E_1$$

Substitute

$$e_{k+1} = e_k - Df(x_k)^{-1} f(x_k)$$

$$= \cancel{e_k} - \cancel{e_k} + Df(x^{(k)})^{-1} E_1 =$$

$$\| e_{k+1} \|_\infty \leq \| Df(x^{(k)})^{-1} E_1 \|_\infty \leq \| Df(x^{(k)})^{-1} \|_\infty \| E_1 \|_\infty$$

Since

$$\max |E_1| \leq \frac{1}{2} A n^2 \| \xi - x^{(k)} \|_\infty^2 = \frac{1}{2} A n^2 \| e_k \|_\infty^2$$

over
all component

$$\| e_{k+1} \|_\infty \leq \underbrace{\| Df(x^{(k)})^{-1} \|_\infty \frac{1}{2} A n^2}\, \| e_k \|_\infty^2$$

Let $\quad C = \max \left\{ \| Df(x^{(k)})^{-1} \|_\infty : x \in B_\varepsilon(\xi) \right\}$

note this maximum exists because $B_\varepsilon(\xi)$ is closed and bounded and $Df(x^{(k)})^{-1}$ is continuous on $B_\varepsilon(\xi)$ be how we chose $\varepsilon$.

Since $x^{(k)} \in B_\varepsilon(\xi)$ then $\| Df(x^{(k)})^{-1} \|_\infty \leq C$.

$$\| e_{k+1} \|_\infty \leq M \| e_k \|_\infty^2 \quad \text{where} \quad M = C \tfrac{1}{2} A n^2$$

In summary... The same proof as in the case of one non linear equation, but Taylor's theorem is more complicated.

# Chapter 5: Eigenvalues and Eigenvectors

Spectral Theorem

**Theorem 5.1** *Suppose that $A \in \mathbb{R}^{n\times n}_{\text{sym}}$; then, the following statements are valid.* ← Symmetric matrices...

(i) *There exist $n$ linearly independent eigenvectors $\boldsymbol{x}^{(i)} \in \mathbb{R}^n$ and corresponding eigenvalues $\lambda_i \in \mathbb{R}$ such that $A\boldsymbol{x}^{(i)} = \lambda_i \boldsymbol{x}^{(i)}$ for all $i = 1, 2, \ldots, n$.*

(ii) *The function*

$$\lambda \mapsto \det(A - \lambda I) \tag{5.2}$$

*is a polynomial of degree $n$ with leading term $(-1)^n\lambda^n$, called the* **characteristic polynomial of** *$A$. The eigenvalues of $A$ are the zeros of the characteristic polynomial.*

(iii) *If the eigenvalues $\lambda_i$ and $\lambda_j$ of $A$ are distinct, then the corresponding eigenvectors $\boldsymbol{x}^{(i)}$ and $\boldsymbol{x}^{(j)}$ are orthogonal in $\mathbb{R}^n$, i.e.,*

$$\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(j)} = 0 \qquad \text{if } \lambda_i \neq \lambda_j\,, \qquad i, j \in \{1, 2, \ldots, n\}\,.$$

(iv) *If $\lambda_i$ is a root of multiplicity $m$ of (5.2), then there is a linear subspace in $\mathbb{R}^n$ of dimension $m$, spanned by $m$ mutually orthogonal eigenvectors associated with the eigenvalue $\lambda_i$.*

(v) *Suppose that each of the eigenvectors $\boldsymbol{x}^{(i)}$ of $A$ is* **normalised**, *in other words, $\boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{x}^{(i)} = 1$ for $i = 1, 2, \ldots, n$, and let $X$ denote the square matrix whose columns are the normalised (orthogonal) eigenvectors; then, the matrix $\Lambda = X^{\mathrm{T}}AX$ is diagonal, and the diagonal elements of $\Lambda$ are the eigenvalues of $A$.*

(vi) *Let $Q \in \mathbb{R}^{n\times n}$ be an orthogonal matrix and define $B \in \mathbb{R}^{n\times n}_{\text{sym}}$ by $B = Q^{\mathrm{T}}AQ$; then, $\det(B - \lambda I) = \det(A - \lambda I)$ for each $\lambda \in \mathbb{R}$. The eigenvalues of $B$ are the same as the eigenvalues of $A$, and the eigenvectors of $B$ are the vectors $Q^{\mathrm{T}}\boldsymbol{x}^{(i)}$, $i = 1, 2, \ldots, n$.*

(vii) *Any vector $\boldsymbol{v} \in \mathbb{R}^n$ can be expressed as a linear combination of the (ortho)normalised eigenvectors $\boldsymbol{x}^{(i)}$, $i = 1, 2, \ldots, n$, of $A$, i.e.,*

$$\boldsymbol{v} = \sum_{i=1}^{n} \alpha_i \boldsymbol{x}^{(i)}, \qquad \alpha_i = \boldsymbol{x}^{(i)\mathrm{T}}\boldsymbol{v}\,.$$

(viii) *The trace of $A$, $\mathrm{Trace}(A) = \sum_{i=1}^{n} a_{ii}$, is equal to the sum of the eigenvalues of $A$.*

Algorithm for finding eigenvalues and eigenvector...

Recall power method: $B = A^T A$

$$y_{k+1} = B x_k$$
$$x_{k+1} = y_{k+1} / \| y_{k+1} \|$$

largest eigenvalue $\approx y_{k+1} \cdot x_k$

eigenvector $\approx x_k$

already know one way to find the largest eigenvalue.

This worked in part because all eigenvalues of $A^T A$ are positive.
If they aren't then there could be two different eigenvalues $\lambda_1 \neq \lambda_2$ such that $|\lambda_1| = |\lambda_2|$ and in this case the method doesn't always work.