Convergence of 2-Stage QR

This QR iteration will converge (with real shifts) to an 'almost' upper triangular matrix, $\hat{U}$,

$$\hat{U} = \begin{bmatrix} \times & \times & \times & \times & \times & \cdots & \times & \times & \times \\ \times & \times & \times & \times & \times & \cdots & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \cdots & \times & \times & \times \\ 0 & 0 & \times & \times & \times & \cdots & \times & \times & \times \\ 0 & 0 & 0 & 0 & \times & \cdots & \times & \times & \times \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \times & \times \\ 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \times & \times \end{bmatrix}.$$

Note that $\hat{U}$ is an upper Hessenberg matrix with every second element of the subdiagonal zero. This implies (from linear algebra) that the eigenvalues of $\hat{U}$ are the union of the eigenvalues of all the diagonal blocks (2×2 and 1×1 blocks).

Nonlinear Systems and Optimization

Nonlinear Systems: – the scalar case

Given $f(x), \;\; f : \Re \to \Re$, find a real root (or zero), α, such that $f(\alpha) = 0$.

We have to be satisfied with an $\bar{\alpha}$ in our FP system, such that $|\bar{\alpha} - \alpha|$ is small.

We will consider, as an example the function $f(x) = x^3 - x - 1$ which has one root in the interval $[0, 2]$.

Recall **Newton's Method for scalars**: Given real numbers $x_0, \; a, b$ and $f(x) \in C^1[a, b]$ with $x_0 \in [a, b]$,

-<u>for</u> $r = 1, 2 \cdots$ <u>until</u> satisfied <u>do</u>:

$$x_{r+1} = x_r - \frac{f(x_r)}{f'(x_r)}$$

-<u>end</u>

After 4 iterations, with $x_0 = 1$ we obtain

$x_4 = 1.32 \cdots$, and $|f(x_4)| \approx 10^{-6}$.

Analysis of Scalar Methods

- Definition: A sequence x_r converges to α iff $|x_r - \alpha| \to 0$ as $r \to \infty$.
- The continuity of $f(x)$ implies
  $x_r \to \alpha \Leftrightarrow |x_r - \alpha| \to 0 \Rightarrow |f(x_r) - f(\alpha)| = |f(x_r)| \to 0$ as $r \to \infty$.
  We can monitor $|f(x_r)|$ on each iteration but how can we recognize $|f(x_r)| \to 0$ and more importantly, how do we recognize that $|x_N - \alpha|$ is small?
- Definition: If $x_r \to \alpha$ and $\rho \geq 1$ is the largest real number such that,

$$\lim_{r \to \infty} \frac{|x_{r+1} - \alpha|}{|x_r - \alpha|^{\rho}} \leq C \neq 0,$$

  for some $C > 0$, then the convergence is order ρ.
  (Note that in this case, if x_r is accurate to k digits then x_{r+1} can be expected to be accurate to $\rho\, k$ digits.)

Newton's Method for Scalars (cont)

Note:

- Newton's method may not converge (the $x_r's$ may $\to \pm\infty$).
- The $|f(x_r)|$ may not decrease (as r increases).
- If $f(x) \in C^1[a,b]$ and $|x_0 - \alpha|$ is sufficiently small, then x_r will converge to α.
- If Newtons method converges and $f'(\alpha) \neq 0$, then the order of convergence is 2.

Systems of Nonlinear Equations

- The Basic Problem: find $x \in \Re^n$ such that $F(x) = \underline{0}$ where,

$$\underline{0} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad F(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}.$$

  Note that, in this section, n is the dimension of x (the number of unknowns) and r will be the iteration number.

From Scalars to Systems

- Analogous with the scalar case, the vector sequence, $x^{(0)}, x^{(1)} \cdots x^{(r)}$ converges to $\underline{\alpha} \in \Re^n$ with order p if there exists $c > 0$ such that

$$\lim_{r \to \infty} \frac{\|x^{(r+1)} - \underline{\alpha}\|}{\|x^{(r)} - \underline{\alpha}\|^p} = c.$$

  The three most common values for p are $p = 1$ (linear convergence), $p = 2$ (quadratic convergence) and $p = 3$ (cubic convergence).

- Of the methods you have investigated for scalar problems ($n = 1$) only Newtons method extends directly to higher dimensional problems ($n > 1$).

NM for Systems

Given an initial guess, $x^{(0)} \in \Re^n$ define $x^{(r)}$ for $r = 0, 1 \cdots$ by solving the linear system,

$$\frac{\partial F}{\partial x}\Big|_{x=x^{(r)}} (x^{(r+1)} - x^{(r)}) = -F(x^{(r)}),$$

or with $\Delta^r = x^{(r+1)} - x^{(r)}$, (the Newton Correction),

$$W_r \Delta^r = -F(x^{(r)}),$$

where $W_r = \frac{\partial F}{\partial x}|_{x=x^{(r)}}$ is the $n \times n$ matrix whose $(i\ j)^{th}$ element is $(\frac{\partial f_i}{\partial x_j})$.

Observations for NM for Systems

- One can re-write this equation as,

$$x^{(r+1)} = x^{(r)} - W_r^{-1} F(x^{(r)}).$$

(Analogous to the scalar case, $x^{r+1} = x^r - f(x^r)/f'(x^r)$.)

- The matrix W_r must be recomputed and a new LU or QR decomposition computed on each iteration.
- If $\frac{\partial F}{\partial x}|_{x=\underline{\alpha}}$ is nonsingular, then if the iterates converge, the convergence is order 2 or quadratic.

Reducing the Cost of NM

Approximate versions of Newtons Method have been developed for systems of nonlinear equations that avoid the $O(n^3)$ flops per iteration or that are more efficient to implement for other reasons.

- If we hold an approximation to W_r constant for several iterations we have a Modified Newton method. For example one can use $W_r = W_1$ for all r or one can re-evaluate W once every k iterations. In either case one looses quadratic convergence.
- If $\frac{\partial F}{\partial x}$ is difficult to compute we can use divided differences to define a Quasi Newton method,

$$\frac{\partial f_i}{\partial x_j} \approx \frac{f_i(x + \delta e_j) - f_i(x)}{\delta},$$

where $\delta \approx \sqrt{\mu}$.

Approximate Newton (Systems)

- We can approximate W_r by a 'nearby' matrix with special structure. This is called pre-conditioning and can involve approximating W_r by a diagonal, banded, or triangular matrix.
- If the approximation $W_r \approx I$ is used the method is called Functional iteration.

These approximate versions of Newtons method often work in special cases and can be readily analysed and justified only for these cases. In the 'very' special case that $F(x)$ is linear, that is, $F(x) = Ax - b$, Newtons method will converge in one iteration since,

$$
\begin{aligned}
\Delta^r &= (x^{(r+1)} - x^{(r)}) \;=\; -W_r^{-1} F(x^{(r)}) \\
&= -A^{-1}(Ax^{(r)} - b) \\
&= A^{-1}b - x^{(r)} \\
&= \alpha - x^{(r)}
\end{aligned}
$$

Summary of NM for Systems

We have shown that the various versions of Newtons Method that are used in practice can be viewed as,

-guess $x^{(0)} = (x_1^{(0)}, x_2^{(0)} \cdots x_n^{(0)})^T$
-<u>for</u> $r = 0, 1 \cdots$ <u>until</u> satisfied <u>do</u>:
  -Solve $W_r \Delta^r = -F(x^{(r)})$
  $x^{(r+1)} = x^{(r)} + \Delta^r$
-<u>end</u>

where $W_r \approx \left(\frac{\partial F}{\partial x}\right)\big|_{x=x^{(r)}}$.

For systems of equations, as for the scalar case, if we observe convergence it is usually very rapid but we need an accurate initial guess to ensure convergence.

Optimization Problems

- A special case of nonlinear systems are optimization problems which arise in a wide variety of application areas. They are usually of the form, Find $x \in \Re^n$ such that $f(x)$ is a minimum (or max).

$$f : \Re^n \rightarrow \Re.$$

- From calculus we know that a vector x is a minimum (or max) when,

$$\nabla f(x) \equiv \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = 0.$$

- $\nabla f(x)$ is called the gradient of f.

WARNING: In this section $f(x)$ and $g(x)$ have a different meaning than in the previous section.