Nonlinear Equations

University of Illinois at Urbana-Champaign

1 These slides have been drafted by Edgar Solomonik as lecture templates and supplementary material for the book “Scientific Computing: An Introductory Survey” by Michael T. Heath (slides).
Solving Nonlinear Equations

- Solving (systems of) nonlinear equations corresponds to root finding:

- Solving nonlinear equations has many applications:
Solving Nonlinear Equations

Main algorithmic approach: find successive roots of local linear approximations of $f$: 
Nonexistence and Nonuniqueness of Solutions

- Solutions do not generally exist and are not generally unique, even in the univariate case:

- Solutions in the multivariate case correspond to intersections of hypersurfaces:
Conditions for Existence of Solution

- Intermediate value theorem for univariate problems:

- A function has a unique fixed point \( g(x^*) = x^* \) in a given closed domain if it is contractive and contained in that domain,

\[
\|g(x) - g(z)\| \leq \gamma \|x - z\|, \quad 0 \leq \gamma < 1
\]
Conditioning of Nonlinear Equations

- Generally, we take interest in the absolute rather than relative conditioning of solving $f(x) = 0$:

  - The \textit{absolute condition number} of finding a root $x^*$ of $f$ is $1/|f'(x^*)|$ and for a root $x^*$ of $f$ it is $||J^{-1}_f(x^*)||$:

  - The \textit{absolute condition number} of finding a root $x^*$ of $f$ is $1/|f'(x^*)|$ and for a root $x^*$ of $f$ it is $||J^{-1}_f(x^*)||$:
Multiple Roots and Degeneracy

- If $x^*$ is a root of $f$ with multiplicity $m$, its $m - 1$ derivatives are also zero at $x^*$,

$$f(x^*) = f'(x^*) = f''(x^*) = \cdots = f^{(m-1)}(x^*) = 0.$$ 

- Increased multiplicity affects conditioning and convergence:
Bisection Algorithm

Assume we know the desired root exists in a bracket \([a, b]\) and 
\[\text{sign}(f(a)) \neq \text{sign}(f(b))\]:

Bisection subdivides the interval by a factor of two at each step by considering \(f(c_k)\) at 
\[c_k = (a_k + b_k)/2\]:
Convergence of Fixed Point Iteration

Fixed point iteration: $x_{k+1} = g(x_k)$ is locally linearly convergent to fixed point $x^*$ if $g$ is continuously differentiable near $x^*$ and $|g'(x^*)| < 1$:

- It is quadratically convergent if $g$ is twice continuously differentiable and $g'(x^*) = 0$:
Newton’s Method

- Newton’s method is derived from a Taylor series expansion of $f$ at $x_k$:

- Newton’s method is \textit{quadratically convergent} if started sufficiently close to $x^*$ so long as $f'(x^*) \neq 0$ and $f$ is twice continuously differentiable in the neighborhood of $x^*$.
Secant Method

- The *Secant method* approximates $f'(x_k) \approx \frac{f(x_k) - f(x_{k-1})}{x_k - x_{k-1}}$:

- The convergence of the Secant method is *superlinear* but not quadratic:
Nonlinear Tangential Interpolants

- Secant method uses a linear interpolant based on points $f(x_k), f(x_{k-1})$, could use more points and higher-order interpolant:

- Quadratic interpolation (Muller’s method) can achieve a convergence order of $r \approx 1.84$:

- Inverse quadratic interpolation resolves the problem of nonexistence/nonuniqueness of roots of polynomial interpolants:
Achieving Global Convergence

► Hybrid bisection/Newton methods:

► Bounded (damped) step-size:
Systems of Nonlinear Equations

Given \( f(x) = [f_1(x) \cdots f_m(x)]^T \) for \( x \in \mathbb{R}^n \), seek \( x^* \) so that \( f(x^*) = 0 \)

At a particular point \( x \), the **Jacobian** of \( f \), describes how \( f \) changes in a given direction of change in \( x \),

\[
J_f(x) = \begin{bmatrix}
\frac{df_1}{dx_1}(x) & \cdots & \frac{df_1}{dx_n}(x) \\
\vdots & \ddots & \vdots \\
\frac{df_m}{dx_1}(x) & \cdots & \frac{df_m}{dx_n}(x)
\end{bmatrix}
\]
Multivariate Newton Iteration

- Fixed-point iteration $x_{k+1} = g(x_k)$ achieves local convergence if (in addition to contraints on differentiability of $g$) we have $|\lambda_{\text{max}}(J_g(x^*))| < 1$ and quadratic convergence if $J_g(x^*) = O$: 

**Demo:** Newton’s method in $n$ dimensions
Multidimensional Newton’s Method

- Newton’s method corresponds to the fixed-point iteration

\[ g(x) = x - J^{-1}_f(x) f(x) \]

- Quadratic convergence is achieved when the Jacobian of a fixed-point iteration is zero at the solution, which is true for Newton’s method:
Estimating the Jacobian using Finite Differences

To obtain $J_f(x_k)$ at iteration $k$, can use finite differences:

- $n + 1$ function evaluations are needed: $f(x)$ and $f(x + he_i), \forall i \in \{1, \ldots, n\}$, which correspond to $m(n + 1)$ scalar function evaluations if $J_f(x_k) \in \mathbb{R}^{m \times n}$.
Cost of Multivariate Newton Iteration

- What is the cost of solving $J_f(x_k)s_k = f(x_k)$?

- What is the cost of Newton’s iteration overall?
Quasi-Newton Methods

In solving a nonlinear equation, seek approximate Jacobian $J_f(x_k)$ for each $x_k$

- Find $B_{k+1} = B_k + \delta B_k \approx J_f(x_{k+1})$, so as to approximate secant equation

$$B_{k+1}(x_{k+1} - x_k) = f(x_{k+1}) - f(x_k)$$


- **Broyden’s method** solves the secant equation and minimizes $||\delta B_k||_F$:

$$\delta B_k = \frac{\delta f - B_k \delta x}{||\delta x||^2} \delta x^T$$
Safeguarding Methods

- Can dampen step-size to improve reliability of Newton or Broyden iteration:

- *Trust region methods* provide general step-size control: