Root-finding for Nonlinear Equations

Dangxing Chen

Duke Kunshan University

Many applications

- Physics, e.g., Navier-Stokes equation
- Chemistry, e.g.,Density functional theory
- Biology, e.g., Molecular dynamics
- Finance, e.g., Simulation
- Data Science, e.g., Machine learning methods
- $\vdots$

N-body simulation

- N-body problem: Predicting the individual motions of a group of celestial objects interacting with each other gravitationally

Figure: `https://insidehpc.com/2015/05/direct-n-body-simulation/`

Diagrams of problem solving

- E.g. N-body simulation
- Identify the problem: Want to predict motions of bodies
- Formulate the problem: Newtonian mechanics

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = -\sum_{j\neq i} \frac{Gm_i m_j(\mathbf{x}_i - \mathbf{x}_j)}{\|\mathbf{x}_i - \mathbf{x}_j\|^3}.$$

- **Question**: How to solve the problem?

Diagrams of problem solving

- E.g. N-body simulation
- Identify the problem: Want to predict motions of bodies
- Formulate the problem: Newtonian mechanics

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = -\sum_{j\neq i} \frac{Gm_i m_j(\mathbf{x}_i - \mathbf{x}_j)}{\|\mathbf{x}_i - \mathbf{x}_j\|^3}.$$

- **Question**: How to solve the problem? Use computer
- Discretize the equation (what is the best way?)
- Efficient method (can we calculate millions or billions of body movements?)
- **Remark**: Numerical analysis allows us to do this simulation!

Root-finding problem

Bisection method

Newton's method

Secant method

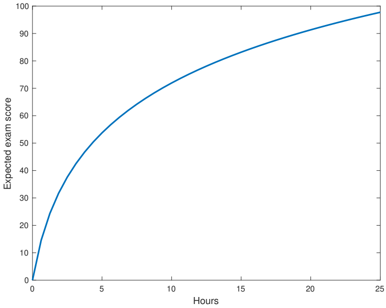
A general theory for one-point iteration methods

System of nonlinear equations

Unconstrained optimization

Introduction

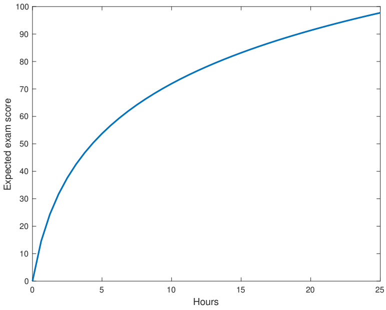
- Expected exam score



- **Question**: How much should I study in order to get 90?

Introduction

- Expected exam score



- **Question**: How much should I study in order to get 90?
- Solve $f(x) = 90$
- In general, solve $f(x) = 0$

Solution method

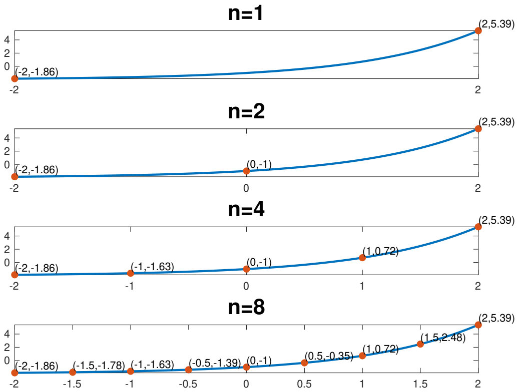
- **Question**: How to solve for $f(x) = 0$?
- If $f(x) = ax + b = 0$, then $x = -\frac{b}{a}$
- If $f(x) = ax^2 + bx + c$, then $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- **Question**: In general case, how to find x?

Solution method

- **Question**: How to solve for $f(x) = 0$?
- If $f(x) = ax + b = 0$, then $x = -\frac{b}{a}$
- If $f(x) = ax^2 + bx + c$, then $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$
- **Question**: In general case, how to find x?
- Suppose the root x^* is in the interval $[a, b]$
- Brute force method: Grid searching
- Calculate $f(x_i)$ and find $\arg\min_{x_i} |f(x_i)|$

Example

- E.g. $f(x) = e^x - 2$ in $[-2, 2]$.
- Root: $x^* = \log(2) \approx 0.69$



Analysis

- **Question**: How to analyze error?

Analysis

- **Question**: How to analyze error?
- **Error**: $e_n = \arg\min_{x_i} |x_i - x^*| \leq \frac{b-a}{n}$
- **Question**: How much points we need such that $e_n < \epsilon$?

Analysis

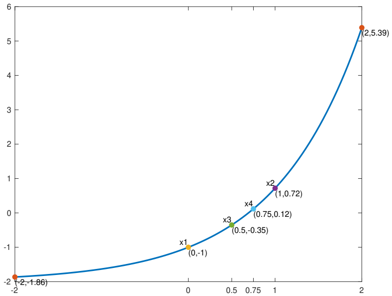
- **Question**: How to analyze error?
- **Error**: $e_n = \arg\min_{x_i} |x_i - x^*| \leq \frac{b-a}{n}$
- **Question**: How much points we need such that $e_n < \epsilon$?
- $\frac{b-a}{n} < \epsilon \Rightarrow n > \frac{b-a}{\epsilon}$
- **Question**: How to know if there is a root in $[a, b]$?

Analysis

- **Question**: How to analyze error?
- **Error**: $e_n = \arg\min_{x_i} |x_i - x^*| \leq \frac{b-a}{n}$
- **Question**: How much points we need such that $e_n < \epsilon$?
- $\frac{b-a}{n} < \epsilon \Rightarrow n > \frac{b-a}{\epsilon}$
- **Question**: How to know if there is a root in $[a, b]$?
- Intermediate value theorem: $f(a)f(b) < 0 \Rightarrow x^* \in (a, b)$
- Assumption of the root-finding problems in general:
  - f is continuous on $[a, b]$
  - $f(a)f(b) < 0$
  - f has exactly one root in $[a, b]$
- **Question**: Can we do better than grid searching?

Bisection method

- **Def: Bisection method**
- Find the midpoint $c = \frac{a+b}{2}$
  - If $f(c) = 0$, done
  - If $f(a)f(c) < 0$, then $x^* \in (a, c)$, let $b = c$
  - If $f(c)f(b) < 0$, then $x^* \in (c, b)$, and $a = c$
  - Repeat
- E.g. $f(x) = e^x - 2$ in $[-2, 2]$.
- Root: $x^* = \log(2) \approx 0.69$



Analysis

- **Question**: How fast does the error decay?

Analysis

- **Question**: How fast does the error decay?
- $e_n \leq \left(\frac{1}{2}\right)^n (b-a)$
- **Question**: How many iterations we need such that $e_n < \epsilon$?

Analysis

- **Question**: How fast does the error decay?
- $e_n \leq \left(\frac{1}{2}\right)^n (b-a)$
- **Question**: How many iterations we need such that $e_n < \epsilon$?
- $\left(\frac{1}{2}\right)^n (b-a) \leq \epsilon \Rightarrow n > \frac{-\log(\epsilon)+\log(b-a)}{\log(2)}$
- E.g. For $a=0$, $b=1$, $\epsilon = 10^{-10}$
  - Grid searching: $n > 10^{10}$
  - Bisection: $n > 33$

Order of convergence

- **Def**: **Iterative method**, generate a sequence of improving approximate solutions
- **Question**: How to measure the interval of convergence and speed of convergence?

Order of convergence

- **Def**: **Iterative method**, generate a sequence of improving approximate solutions
- **Question**: How to measure the interval of convergence and speed of convergence?
- **Def**: A sequence of iterates $\{x_n\}$ is said to converge with **order** $p \geq 1$ to a point x^* if

$$|x^* - x_{n+1}| \leq c|x^* - x_n|^p$$

  for some $c > 0$. If $p = 1$, the sequence is said to **converge linearly** to x^*.
- **Question**: What is order of Bisection?

Order of convergence

- **Def**: **Iterative method**, generate a sequence of improving approximate solutions
- **Question**: How to measure the interval of convergence and speed of convergence?
- **Def**: A sequence of iterates $\{x_n\}$ is said to converge with **order** $p \geq 1$ to a point x^* if

$$|x^* - x_{n+1}| \leq c|x^* - x_n|^p$$

  for some $c > 0$. If $p = 1$, the sequence is said to **converge linearly** to x^*.
- **Question**: What is order of Bisection? Linear with $c = \frac{1}{2}$
- **Question**: Why this measures the convergence?

Order of convergence

- **Def**: **Iterative method**, generate a sequence of improving approximate solutions
- **Question**: How to measure the interval of convergence and speed of convergence?
- **Def**: A sequence of iterates $\{x_n\}$ is said to converge with **order** $p \geq 1$ to a point x^* if

$$|x^* - x_{n+1}| \leq c|x^* - x_n|^p$$

  for some $c > 0$. If $p = 1$, the sequence is said to **converge linearly** to x^*.
- **Question**: What is order of Bisection? Linear with $c = \frac{1}{2}$
- **Question**: Why this measures the convergence?
- Suppose $p = 1$, if $c > 1$, we do not necessarily have convergence
- **Question**: Why this measures the speed of convergence?

Speed of convergence

- Suppose $p = 1$

$$
\begin{aligned}
e_1 &\le ce_0, \\
e_2 &\le ce_1 \le c^2 e_0, \\
e_3 &\le ce_2 \le c^2 e_1 \le c^3 e_0, \\
&\vdots \\
e_n &\le c^n e_0.
\end{aligned}
$$

- Suppose $p = 2$

$$
\begin{aligned}
e_1 &\le ce_0^2, \\
e_2 &\le ce_1^2 \le c^3 e_0^4, \\
e_3 &\le ce_2^2 \le c^3 e_1^4 \le c^7 e_0^8, \\
&\vdots \\
e_n &\le c^{2^n-1} e_0^{2^n}.
\end{aligned}
$$

Speed of convergence

- For $p = 1$, $e_n \leq c^n e_0$
- For $p = 2$, $e_n \leq c^{2^n - 1} e_0^{2^n}$
- **Question**: For the same c, how much faster for $p = 2$?

Speed of convergence

- For $p = 1$, $e_n \le c^n e_0$
- For $p = 2$, $e_n \le c^{2^n-1} e_0^{2^n}$
- **Question**: For the same c, how much faster for $p = 2$?
- For $p = 1$, let $e_m \le c^m e_0$
- For $p = 2$, let $e_n \le c^{2^n-1} e_0^{2^n}$
- Let $c^m e_0 = c^{2^n-1} e_0^{2^n}$, then $m = \frac{(2^n-1)\log(c)+(2^n-1)\log(e_0)}{\log(c)}$
- E.g., if $e_0 = c$, then $m = 2^{n+1} - 2$
- E.g., for $n = 5$, $m = 62$
- **Question**: Can we improve the Bisection's method to second order?

Taylor series

- **Mean Value Theorem** (MVT): Let $f(x)$ be continuous for $a \leq x \leq b$, and let it be differentiable for $a < x < b$. Then there is at least one point ξ in (a, b) for which

$$f(b) - f(a) = f'(\xi)(b - a).$$

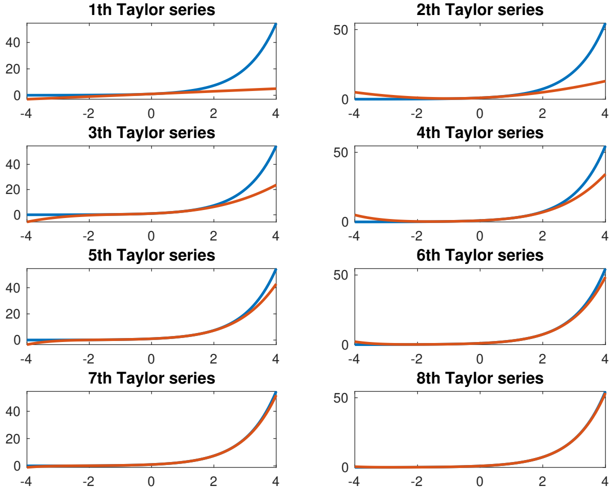
- **Taylor's Theorem**: Let $f(x)$ be infinitely continuous differentiable on $[a, b]$, and let $x, x_0 \in [a, b]$. Then

$$\begin{aligned} f(x) &= p_n(x) + R_{n+1}(x), \\ p_n(x) &= f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0), \\ R_{n+1}(x) &= \frac{(x - x_0)^{n+1}}{(n+1)!} f^{(n+1)}(\xi), \end{aligned}$$

for some ξ between x_0 and x.

Example

- E.g., $f(x) = e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} = 1 + x + \frac{x^2}{2} + \frac{x^3}{6} + \dots$



Taylor series - Motivation

- **Question**: Why Taylor series makes sense?
- n-th order Taylor series at $x = x_0$

$$f(x) \approx p_n(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0)$$

Taylor series - Motivation

- **Question**: Why Taylor series makes sense?
- n-th order Taylor series at $x = x_0$

$$f(x) \approx p_n(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0)$$

- **Goal**: Fit a n-th degree polynomial to $f(x)$ around $x = x_0$
- $p_n(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + \cdots + c_n(x - x_0)^n$

Taylor series - Motivation

- **Question**: Why Taylor series makes sense?
- n-th order Taylor series at $x = x_0$

$$f(x) \approx p_n(x) = f(x_0) + \frac{(x - x_0)}{1!} f'(x_0) + \cdots + \frac{(x - x_0)^n}{n!} f^{(n)}(x_0)$$

- **Goal**: Fit a n-th degree polynomial to $f(x)$ around $x = x_0$
- $p_n(x) = c_0 + c_1(x - x_0) + c_2(x - x_0)^2 + \cdots + c_n(x - x_0)^n$
- Want $f^{(k)}(x_0) = p_n^{(k)}(x_0)$ for $k \le n$
- $f(x_0) = p_n(x_0) = c_0$
- $f'(x_0) = p_n'(x_0) = c_1$
- $f''(x_0) = p_n''(x_0) = 2c_2 \Rightarrow c_2 = \frac{f''(x_0)}{2}$
- $\vdots$

Newton's method

- Suppose we have a good initial guess x_0
- Motivation: Taylor expansion

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(\xi),$$

  where ξ is between x and x_0
- **Question**: What happens if x^* is the true solution?

Newton's method

- Suppose we have a good initial guess x_0
- Motivation: Taylor expansion

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(\xi),$$

  where ξ is between x and x_0
- **Question**: What happens if x^* is the true solution?
- $0 = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^* - x_0)^2}{2} f''(\xi)$
- $x^* = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{(x^* - x_0)^2}{2} \cdot \frac{f''(\xi)}{f'(x_0)}$
- **Question**: How to solve this equation?

Newton's method

- Suppose we have a good initial guess x_0
- Motivation: Taylor expansion

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(\xi),$$

  where ξ is between x and x_0
- **Question**: What happens if x^* is the true solution?
- $0 = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^*-x_0)^2}{2} f''(\xi)$
- $x^* = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{(x^*-x_0)^2}{2} \cdot \frac{f''(\xi)}{f'(x_0)}$
- **Question**: How to solve this equation?
- $x^* \approx x_0 - \frac{f(x_0)}{f'(x_0)}$
- **Question**: Should we just stop here?

Newton's method

- Suppose we have a good initial guess x_0
- Motivation: Taylor expansion

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(\xi),$$

  where ξ is between x and x_0
- **Question**: What happens if x^* is the true solution?
- $0 = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^*-x_0)^2}{2} f''(\xi)$
- $x^* = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{(x^*-x_0)^2}{2} \cdot \frac{f''(\xi)}{f'(x_0)}$
- **Question**: How to solve this equation?
- $x^* \approx x_0 - \frac{f(x_0)}{f'(x_0)}$
- **Question**: Should we just stop here?
- Iterate: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Question**: Can this converge?

Newton's method

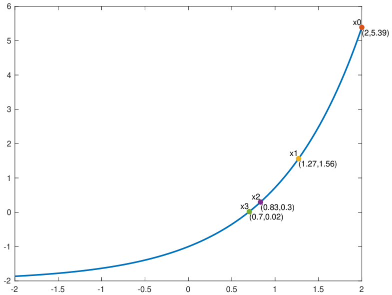
- Suppose we have a good initial guess x_0
- Motivation: Taylor expansion

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \frac{(x - x_0)^2}{2} f''(\xi),$$

  where ξ is between x and x_0
- **Question**: What happens if x^* is the true solution?
- $0 = f(x_0) + (x^* - x_0)f'(x_0) + \frac{(x^* - x_0)^2}{2} f''(\xi)$
- $x^* = x_0 - \frac{f(x_0)}{f'(x_0)} - \frac{(x^* - x_0)^2}{2} \cdot \frac{f''(\xi)}{f'(x_0)}$
- **Question**: How to solve this equation?
- $x^* \approx x_0 - \frac{f(x_0)}{f'(x_0)}$
- **Question**: Should we just stop here?
- Iterate: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Question**: Can this converge?
- Each time, error $\frac{(x^* - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)}$ becomes smaller

Example

- E.g. $f(x) = e^x - 2$ in $[-2, 2]$.
- Root: $x^* = \log(2) \approx 0.69$
- $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{e^{x_n} - 2}{e^{x_n}}$



Convergence analysis

- **Question**: How fast does the Newton's method converges?

Convergence analysis

- **Question**: How fast does the Newton's method converges?
- $e_{n+1} = \left| \frac{(x^* - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)} \right|$
- Suppose $\left| \frac{1}{2} \frac{f''(\xi_n)}{f'(x_n)} \right| \leq M$
- $e_{n+1} \leq M e_n^2$
- **Question**: What is the order of the Newton's method?

Convergence analysis

- **Question**: How fast does the Newton's method converges?
- $e_{n+1} = \left| \frac{(x^* - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)} \right|$
- Suppose $\left| \frac{1}{2} \frac{f''(\xi_n)}{f'(x_n)} \right| \leq M$
- $e_{n+1} \leq M e_n^2$
- **Question**: What is the order of the Newton's method?
- Second order
- **Question**: $\lim_{n \to \infty} \frac{e_{n+1}}{e_n^2} = ?$

Convergence analysis

- **Question**: How fast does the Newton's method converges?
- $e_{n+1} = \left| \frac{(x^* - x_n)^2}{2} \frac{f''(\xi_n)}{f'(x_n)} \right|$
- Suppose $\left| \frac{1}{2} \frac{f''(\xi_n)}{f'(x_n)} \right| \leq M$
- $e_{n+1} \leq M e_n^2$
- **Question**: What is the order of the Newton's method?
- Second order
- **Question**: $\lim_{n\to\infty} \frac{e_{n+1}}{e_n^2} = ?$
- **THM**: $\lim_{n\to\infty} \frac{e_{n+1}}{e_n^2} = \left| \frac{f''(x^*)}{2f'(x^*)} \right|$

Convergence analysis - continued

- $e_{n+1} \leq M e_n^2$
- **Question**: Does this necessarily always converge?

Convergence analysis - continued

- $e_{n+1} \leq M e_n^2$
- **Question**: Does this necessarily always converge?
- E.g., $f(x) = x^3 - 2x + 2$
- $f'(x) = 3x^2 - 2$
- $x_{n+1} = x_n - \frac{x_n^3 - 2x_n + 2}{3x_n^2 - 2}$.
- Root: $x^* \approx -1.77$
- $x_0 = 0$
- $1 \to 0 \to 1 \to 0 \to 1$
- We need a good initial guess

Convergence (failure)

- $f(x) = \sqrt[3]{x}$ with root $x^* = 0$
- Newton:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^{\frac{1}{3}}}{\frac{1}{3}x_n^{-\frac{2}{3}}} = x_n - 3x_n = -2x_n.$$

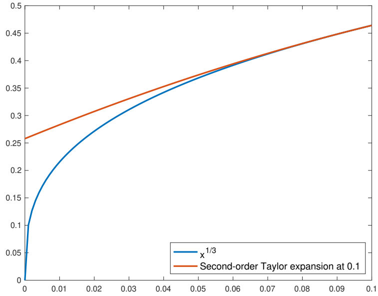
- Never converge
- **Question**: Why?

Convergence (failure)

- $f(x) = \sqrt[3]{x}$ with root $x^* = 0$
- Newton:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{x_n^{\frac{1}{3}}}{\frac{1}{3}x_n^{-\frac{2}{3}}} = x_n - 3x_n = -2x_n.$$

- Never converge
- **Question**: Why?
- Taylor series doesn't work well near the origin



Secant method

- Newton's method: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Question**: What if we don't know $f'(x)$?

Secant method

- Newton's method: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Question**: What if we don't know $f'(x)$?
- Numerical differentiation
- $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$
- $f'(x) \approx \frac{f(x+h)-f(x)}{h}$
- **Question**: Is there a way to save some computational efforts?

Secant method

- Newton's method: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Question**: What if we don't know $f'(x)$?
- Numerical differentiation
- $f'(x) = \lim_{h \to 0} \frac{f(x+h)-f(x)}{h}$
- $f'(x) \approx \frac{f(x+h)-f(x)}{h}$
- **Question**: Is there a way to save some computational efforts?
- Let $h = x_{n-1} - x_n$
- $f'(x_n) \approx \frac{f(x_{n-1})-f(x_n)}{x_{n-1}-x_n} = \frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$
- **Secant method**:

$$x_{n+1} = x_n - f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}.$$

- **Question**: What is the order of the Secant method?

Convergence analysis

- Notations:
  - $f[x_{n-1}, x_n] = \frac{f(x_n) - f(x_{n-1})}{x_n - x_{n-1}}$
  - $f[x_{n-1}, x_n, x_{n+1}] = \frac{f[x_n, x_{n+1}] - f[x_{n-1}, x_n]}{x_{n+1} - x_{n-1}}$
  - $\vdots$
- Error analysis:

$$x^* - x_{n+1} = x^* - x_n + f(x_n) \cdot \frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}.$$

- By algebraic manipulation (HW)

$$x^* - x_{n+1} = -(x^* - x_{n-1})(x^* - x_n)\frac{f[x_{n-1}, x_n, x^*]}{f[x_{n-1}, x_n]}.$$

- Use MVT

$$x^* - x_{n+1} = -(x^* - x_{n-1})(x^* - x_n) \cdot \frac{f''(\zeta_n)}{2f'(\xi_n)}.$$

- **Question**: What is the order?

Convergence analysis

- Let $M = \frac{\max_x |f''(x)|}{2\min_x |f'(x)|}$
- $e_2 \leq M e_1 e_0$
- **Question**: How to get a better recurrence relationship?

Convergence analysis

- Let $M = \frac{\max_x |f''(x)|}{2\min_x |f'(x)|}$
- $e_2 \leq M e_1 e_0$
- **Question**: How to get a better recurrence relationship?
- $M e_2 \leq M e_1 \cdot M e_0$
- **Question**: When does this converge (which also simplify the calculation)?

Convergence analysis

- Let $M = \frac{\max_x |f''(x)|}{2\min_x |f'(x)|}$
- $e_2 \leq M e_1 e_0$
- **Question**: How to get a better recurrence relationship?
- $Me_2 \leq Me_1 \cdot Me_0$
- **Question**: When does this converge (which also simplify the calculation)?
- $\delta = \max\{Me_0, Me_1\} < 1$
- $Me_2 \leq \delta^2$
- $Me_3 \leq Me_2 \cdot Me_1 \leq \delta^3$
- $Me_4 \leq Me_3 \cdot Me_2 \leq \delta^5$
- $Me_n \leq \delta^{q_n}$
- **Question**: $q_n =?$

Convergence analysis

- Let $M = \frac{\max_x |f''(x)|}{2\min_x |f'(x)|}$
- $e_2 \le M e_1 e_0$
- **Question**: How to get a better recurrence relationship?
- $M e_2 \le M e_1 \cdot M e_0$
- **Question**: When does this converge (which also simplify the calculation)?
- $\delta = \max\{M e_0, M e_1\} < 1$
- $M e_2 \le \delta^2$
- $M e_3 \le M e_2 \cdot M e_1 \le \delta^3$
- $M e_4 \le M e_3 \cdot M e_2 \le \delta^5$
- $M e_n \le \delta^{q_n}$
- **Question**: $q_n = ?$
- **Fibonacci sequence**: $q_{n+1} = q_n + q_{n-1}$ with

Convergence analysis

- Let $M = \frac{\max_x |f''(x)|}{2\min_x |f'(x)|}$
- $e_2 \leq M e_1 e_0$
- **Question**: How to get a better recurrence relationship?
- $Me_2 \leq Me_1 \cdot Me_0$
- **Question**: When does this converge (which also simplify the calculation)?
- $\delta = \max\{Me_0, Me_1\} < 1$
- $Me_2 \leq \delta^2$
- $Me_3 \leq Me_2 \cdot Me_1 \leq \delta^3$
- $Me_4 \leq Me_3 \cdot Me_2 \leq \delta^5$
- $Me_n \leq \delta^{q_n}$
- **Question**: $q_n = ?$
- **Fibonacci sequence**: $q_{n+1} = q_n + q_{n-1}$ with $q_0 = q_1 = 1$
- **Question**: How to solve this?

Solution to the sequence

- Matrix equation:

$$\begin{bmatrix} q_n \\ q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \end{bmatrix}$$

- $\mathbf{q}_{n+1} = \mathbf{A}\mathbf{q}_n$
- **Question**: What is the relationship between $\mathbf{q}_{n+1}$ and $\mathbf{q}_1$?

Solution to the sequence

- Matrix equation:

$$\begin{bmatrix} q_n \\ q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \end{bmatrix}$$

- $\mathbf{q}_{n+1} = \mathbf{A}\mathbf{q}_n$
- **Question**: What is the relationship between $\mathbf{q}_{n+1}$ and $\mathbf{q}_1$?
- $\mathbf{q}_{n+1} = \mathbf{A}^n\mathbf{q}_1$
- **Question**: How to calculate $\mathbf{A}^n$?

Solution to the sequence

- Matrix equation:

$$\begin{bmatrix} q_n \\ q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \end{bmatrix}$$

- $\mathbf{q}_{n+1} = \mathbf{A}\mathbf{q}_n$
- **Question**: What is the relationship between $\mathbf{q}_{n+1}$ and $\mathbf{q}_1$?
- $\mathbf{q}_{n+1} = \mathbf{A}^n\mathbf{q}_1$
- **Question**: How to calculate $\mathbf{A}^n$?
- Use eigen-decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, then $\mathbf{A}^n = \mathbf{V}\mathbf{D}^n\mathbf{V}^{-1}$
- **Question**: How to calculate the eigen-decomposition?

Solution to the sequence

- Matrix equation:

$$\begin{bmatrix} q_n \\ q_{n+1} \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} q_{n-1} \\ q_n \end{bmatrix}$$

- $\mathbf{q}_{n+1} = \mathbf{A}\mathbf{q}_n$
- **Question**: What is the relationship between $\mathbf{q}_{n+1}$ and $\mathbf{q}_1$?
- $\mathbf{q}_{n+1} = \mathbf{A}^n\mathbf{q}_1$
- **Question**: How to calculate $\mathbf{A}^n$?
- Use eigen-decomposition $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$, then $\mathbf{A}^n = \mathbf{V}\mathbf{D}^n\mathbf{V}^{-1}$
- **Question**: How to calculate the eigen-decomposition?
- Solve for eigenvalues: $\det(\mathbf{A} - \lambda\mathbf{I}) = 0$
- $\begin{vmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = (-\lambda)(1-\lambda) - 1 = 0 \Rightarrow \lambda = \frac{1 \pm \sqrt{5}}{2}$

Eigen-decomposition

- Solve for eigenvectors: $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$
- $\begin{bmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} -\lambda v_1 + v_2 & = 0, \\ v_1 + (1-\lambda) v_2 & = 0. \end{cases}$
- Reduces to $-\lambda v_1 + v_2 = 0$
- $\mathbf{v} = \begin{bmatrix} 1 \\ \lambda \end{bmatrix}$
- $\mathbf{V} = \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}$ and $\mathbf{V}^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} \frac{\sqrt{5}-1}{2} & 1 \\ \frac{\sqrt{5}+1}{2} & -1 \end{bmatrix}$
- $\mathbf{A} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}-1}{2} & 1 \\ \frac{\sqrt{5}+1}{2} & -1 \end{bmatrix}$
- **Question**: Can we simplify the calculation?

Eigen-decomposition

- Solve for eigenvectors: $(\mathbf{A} - \lambda \mathbf{I})\mathbf{v} = \mathbf{0}$
- $\begin{bmatrix} -\lambda & 1 \\ 1 & 1-\lambda \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow \begin{cases} -\lambda v_1 + v_2 & = 0, \\ v_1 + (1-\lambda) v_2 & = 0. \end{cases}$
- Reduces to $-\lambda v_1 + v_2 = 0$
- $\mathbf{v} = \begin{bmatrix} 1 \\ \lambda \end{bmatrix}$
- $\mathbf{V} = \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix}$ and $\mathbf{V}^{-1} = \frac{1}{\sqrt{5}} \begin{bmatrix} \frac{\sqrt{5}-1}{2} & 1 \\ \frac{\sqrt{5}+1}{2} & -1 \end{bmatrix}$
- $\mathbf{A} = \frac{1}{\sqrt{5}} \begin{bmatrix} 1 & 1 \\ \frac{1+\sqrt{5}}{2} & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \frac{1+\sqrt{5}}{2} & 0 \\ 0 & \frac{1-\sqrt{5}}{2} \end{bmatrix} \begin{bmatrix} \frac{\sqrt{5}-1}{2} & 1 \\ \frac{\sqrt{5}+1}{2} & -1 \end{bmatrix}$
- **Question**: Can we simplify the calculation?

Calculation

- Write $\mathbf{q}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$
- $\mathbf{A}^n\mathbf{q}_1 = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2$
- **Question**: How to solve for c_i?

Calculation

- Write $\mathbf{q}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$
- $\mathbf{A}^n\mathbf{q}_1 = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2$
- **Question**: How to solve for c_i?
- Solve $\mathbf{V}\mathbf{c} = \mathbf{q}_1$
- $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1 \\ -\lambda_2 \end{bmatrix}$
- $\mathbf{q}_{n+1} = \mathbf{A}^n\mathbf{q}_1 = \frac{1}{\sqrt{5}}\lambda_1^{n+1}\mathbf{v}_1 - \frac{1}{\sqrt{5}}\lambda_2^{n+1}\mathbf{v}_2$
- $q_{n+1} = \frac{1}{\sqrt{5}}\left(\lambda_1^{n+2} - \lambda_2^{n+2}\right)$
- **Question**: What happens for large n?

Calculation

- Write $\mathbf{q}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$
- $\mathbf{A}^n\mathbf{q}_1 = c_1\lambda_1^n\mathbf{v}_1 + c_2\lambda_2^n\mathbf{v}_2$
- **Question**: How to solve for c_i?
- Solve $\mathbf{V}\mathbf{c} = \mathbf{q}_1$
- $\mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} \lambda_1 \\ -\lambda_2 \end{bmatrix}$
- $\mathbf{q}_{n+1} = \mathbf{A}^n\mathbf{q}_1 = \frac{1}{\sqrt{5}}\lambda_1^{n+1}\mathbf{v}_1 - \frac{1}{\sqrt{5}}\lambda_2^{n+1}\mathbf{v}_2$
- $q_{n+1} = \frac{1}{\sqrt{5}}\left(\lambda_1^{n+2} - \lambda_2^{n+2}\right)$
- **Question**: What happens for large n?
- $\lambda_2 = \frac{1-\sqrt{5}}{2} \approx -0.62 < 1$
- $\lim_{n\to\infty} \lambda_2^n = 0$
- $q_n \approx \frac{1}{\sqrt{5}}\lambda_1^{n+1}$ for large n
- $e_n \leq \frac{1}{M}\delta^{q_n}$
- **Question**: What is the order?

Order

- Informal argument
- Want $e_{n+1} \leq c e_n^p$ or $\frac{e_{n+1}}{e_n^p} \leq c$
- $e_n \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+1}}$
- $e_{n+1} \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+2}}$
- Calculate $e_n^{\lambda_1}$ to get ride of n
- So order is $p = \lambda_1$
- $1 < \frac{1+\sqrt{5}}{2} \approx 1.62 < 2$
- **Remark**: Faster than the Bisection but slower than the Newton
- **Question**: $c =?$

Order

- Informal argument
- Want $e_{n+1} \leq c e_n^p$ or $\frac{e_{n+1}}{e_n^p} \leq c$
- $e_n \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+1}}$
- $e_{n+1} \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+2}}$
- Calculate $e_n^{\lambda_1}$ to get ride of n
- So order is $p = \lambda_1$
- $1 < \frac{1+\sqrt{5}}{2} \approx 1.62 < 2$
- **Remark**: Faster than the Bisection but slower than the Newton
- **Question**: $c = ?$
- $M^{\lambda_1 - 1} = M^{\frac{\sqrt{5}-1}{2}}$
- **Question**: $\lim_{n\to\infty} \frac{e_{n+1}}{e_n^p}$?

Order

- Informal argument
- Want $e_{n+1} \leq c e_n^p$ or $\frac{e_{n+1}}{e_n^p} \leq c$
- $e_n \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+1}}$
- $e_{n+1} \leq \frac{1}{M} \delta^{\frac{1}{\sqrt{5}} \lambda_1^{n+2}}$
- Calculate $e_n^{\lambda_1}$ to get ride of n
- So order is $p = \lambda_1$
- $1 < \frac{1+\sqrt{5}}{2} \approx 1.62 < 2$
- **Remark**: Faster than the Bisection but slower than the Newton
- **Question**: $c = ?$
- $M^{\lambda_1 - 1} = M^{\frac{\sqrt{5}-1}{2}}$
- **Question**: $\lim_{n\to\infty} \frac{e_{n+1}}{e_n^p}$?
- $\left| \frac{f''(x^*)}{2f'(x^*)} \right|^{\frac{\sqrt{5}-1}{2}}$

Comparison

- **Question**: What are pros and cons?

Comparison

- **Question**: What are pros and cons?
- Bisection: Stable, but slow
- Newton: Fastest, need good initial condition, need to know derivative
- Secant: Fast, need good initial condition, don't need derivative, computational cost less than the Newton in each step

Numerical calculation of order

- Order: $e_{n+1} \leq c e_n^p$
- **Question**: How to verify this numerically?

Numerical calculation of order

- Order: $e_{n+1} \leq ce_n^p$
- **Question**: How to verify this numerically?
- $\log(e_{n+1}) \approx \log(c) + p\log(e_n)$
- Fit this by a linear function $y = ax + b$
- $a = p$, $b = \log(c)$
- **Question**: In practice, lots of fluctuations, how to fit y?

Numerical calculation of order

- Order: $e_{n+1} \leq c e_n^p$
- **Question**: How to verify this numerically?
- $\log(e_{n+1}) \approx \log(c) + p \log(e_n)$
- Fit this by a linear function $y = ax + b$
- $a = p$, $b = \log(c)$
- **Question**: In practice, lots of fluctuations, how to fit y?
- Use the least-squares method to solve this
- $e_n = \frac{1}{n} \sum_{i=1}^{n} (y_i - (ax_i + b))^2$
- $\min_{a,b} e_n$
- **Question**: How to solve this?

Numerical calculation of order

- Order: $e_{n+1} \leq c e_n^p$
- **Question**: How to verify this numerically?
- $\log(e_{n+1}) \approx \log(c) + p\log(e_n)$
- Fit this by a linear function $y = ax + b$
- $a = p$, $b = \log(c)$
- **Question**: In practice, lots of fluctuations, how to fit y?
- Use the least-squares method to solve this
- $e_n = \frac{1}{n}\sum_{i=1}^{n}(y_i - (ax_i + b))^2$
- $\min_{a,b} e_n$
- **Question**: How to solve this? (HW)
- $\frac{\partial e_n}{\partial a} = 0 = \frac{\partial e_n}{\partial b}$
- $\widehat{a} = \overline{y} - \widehat{b}\overline{x}$
- $\widehat{b} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n}(x_i - \overline{x})^2}$

Fixed-point iteration

- Newton: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Fixed-point iteration**: $x_{n+1} = g(x_n)$
- **Question**: What is the theoretical solution here?

Fixed-point iteration

- Newton: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Fixed-point iteration**: $x_{n+1} = g(x_n)$
- **Question**: What is the theoretical solution here?
- **Def**: $x = g(x)$
- There are many ways to reformulate the root-finding to a fixed-point problem
- E.g. $x^2 - a = 0$

Fixed-point iteration

- Newton: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Fixed-point iteration**: $x_{n+1} = g(x_n)$
- **Question**: What is the theoretical solution here?
- **Def**: $x = g(x)$
- There are many ways to reformulate the root-finding to a fixed-point problem
- E.g. $x^2 - a = 0$
- $x_{n+1} = x_n^2 + x_n - a$
- $x_{n+1} = \frac{a}{x_n}$
- $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$
- **Question**: When does the fixed-point problem has a solution?

Fixed-point iteration

- Newton: $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$
- **Fixed-point iteration**: $x_{n+1} = g(x_n)$
- **Question**: What is the theoretical solution here?
- **Def**: $x = g(x)$
- There are many ways to reformulate the root-finding to a fixed-point problem
- E.g. $x^2 - a = 0$
- $x_{n+1} = x_n^2 + x_n - a$
- $x_{n+1} = \frac{a}{x_n}$
- $x_{n+1} = \frac{1}{2}\left(x_n + \frac{a}{x_n}\right)$
- **Question**: When does the fixed-point problem has a solution?
- **THM**: If $g(x)$ be continuous on $[a, b]$ and $a \leq g(x) \leq b$, then at least one solution.
- By IVT

Some analysis

- Suppose $g(x)$ continuous on $[a, b]$ and $g([a, b]) \subset [a, b]$
- Suppose we further know $0 < \lambda < 1$ s.t. $|g(x) - g(y)| \leq \lambda |x - y|, \forall x, y$
- **Question**: Is the solution unique?

Some analysis

- Suppose $g(x)$ continuous on $[a, b]$ and $g([a, b]) \subset [a, b]$
- Suppose we further know $0 < \lambda < 1$ s.t. $|g(x) - g(y)| \leq \lambda |x - y|, \forall x, y$
- **Question**: Is the solution unique?
- Suppose there are two solutions α and β
- $|\alpha - \beta| = |g(\alpha) - g(\beta)| \leq \lambda |\alpha - \beta|$
- **THM**: Unique solution
- **Question**: What can we say about the convergence?

Some analysis

- Suppose $g(x)$ continuous on $[a, b]$ and $g([a, b]) \subset [a, b]$
- Suppose we further know $0 < \lambda < 1$ s.t. $|g(x) - g(y)| \leq \lambda |x - y|, \forall x, y$
- **Question**: Is the solution unique?
- Suppose there are two solutions α and β
- $|\alpha - \beta| = |g(\alpha) - g(\beta)| \leq \lambda |\alpha - \beta|$
- **THM**: Unique solution
- **Question**: What can we say about the convergence?
- $e_{n+1} = |x^* - x_{n+1}| = |g(x^*) - g(x_n)| \leq \lambda |x^* - x_n| = \lambda e_n$
- **THM**: $e_n \leq \lambda^n e_0$
- $|x^* - x_0| \leq |x^* - x_1| + |x_1 - x_0| \leq \lambda |x^* - x_0| + |x_1 - x_0|$
- $|x^* - x_0| \leq \frac{1}{1-\lambda} |x_1 - x_0|$
- **THM**: $e_n \leq \frac{\lambda^n}{1-\lambda} |x_1 - x_0|$

Analysis - continued

- **Question**: Can we calculate the current error based on λ?

Analysis - continued

- **Question**: Can we calculate the current error based on λ?
- $e_n \leq \frac{1}{1-\lambda}|x_{n+1} - x_n|$
- **Question**: How do we know if such λ exists?

Analysis - continued

- **Question**: Can we calculate the current error based on λ?
- $e_n \leq \frac{1}{1-\lambda}|x_{n+1} - x_n|$
- **Question**: How do we know if such λ exists?
- $g(x) - g(y) = g'(\xi)(x - y)$
- Let $\lambda = \max_{a \leq x \leq b} |g'(x)|$
- **Question**: Checking the whole interval seems require a lot of work. Can we simplify it?

Analysis - continued

- **Question**: Can we calculate the current error based on λ?
- $e_n \leq \frac{1}{1-\lambda}|x_{n+1} - x_n|$
- **Question**: How do we know if such λ exists?
- $g(x) - g(y) = g'(\xi)(x - y)$
- Let $\lambda = \max_{a \leq x \leq b} |g'(x)|$
- **Question**: Checking the whole interval seems require a lot of work. Can we simplify it?
- **THM**: Suppose that $g(x)$ is continuously differentiable in some neighboring interval about x^* with $|g'(x^*)| < 1$. Then the result still holds.
- e.g. $x^2 - 3 = 0$, for $x^* = \sqrt{3}$
- e.g. $g(x) = x^2 + x - 3$, $g'(x^*) = 2\sqrt{3} + 1 > 1$, so diverges
- e.g. $g(x) = \frac{1}{2}\left(x + \frac{3}{x}\right)$, $g'(x^*) = \frac{1}{2}\left(1 - \frac{3}{\sqrt{3}^2}\right) = 0$, so converges
- **Question**: What is the order of the fixed-point iteration here?

Analysis - continued

- **Question**: Can we calculate the current error based on λ?
- $e_n \leq \frac{1}{1-\lambda}|x_{n+1} - x_n|$
- **Question**: How do we know if such λ exists?
- $g(x) - g(y) = g'(\xi)(x - y)$
- Let $\lambda = \max_{a \leq x \leq b} |g'(x)|$
- **Question**: Checking the whole interval seems require a lot of work. Can we simplify it?
- **THM**: Suppose that $g(x)$ is continuously differentiable in some neighboring interval about x^* with $|g'(x^*)| < 1$. Then the result still holds.
- e.g. $x^2 - 3 = 0$, for $x^* = \sqrt{3}$
- e.g. $g(x) = x^2 + x - 3$, $g'(x^*) = 2\sqrt{3} + 1 > 1$, so diverges
- e.g. $g(x) = \frac{1}{2}\left(x + \frac{3}{x}\right)$, $g'(x^*) = \frac{1}{2}\left(1 - \frac{3}{\sqrt{3}^2}\right) = 0$, so converges
- **Question**: What is the order of the fixed-point iteration here? First

Higher-order method

- **Question**: How did Newton get second order?

Higher-order method

- **Question**: How did Newton get second order?
- $g'(x^*) = 0$
- $g(x) = x - \frac{f(x)}{f'(x)}$
- $g'(x) = 1 - \frac{(f'(x))^2 - f(x)f''(x)}{(f'(x))^2}$
- $x_{n+1} = g(x_n) = g(x^*) + (x_n - x^*)g'(x^*) + \frac{(x_n - x^*)^2}{2} g''(\xi_n)$
- $\left|\frac{e_{n+1}}{e_n^2}\right| = \left|\frac{1}{2} g''(\xi_n)\right|$
- **THM**: If $g'(x^*) = g''(x^*) = \cdots = g^{(p-1)}(x^*) = 0$, then p order convergence with

$$\lim_{n\to\infty}\left|\frac{e_{n+1}}{e_n^p}\right| = \left|\frac{g^{(p)}(x^*)}{p!}\right|$$

Function with many variables

- For simplicity, focus on 2-d, $f(x, y)$
- Taylor series in 2-d around (a, b):

$$
\begin{aligned}
f(x,y) \approx f(a,b) &+ (x-a) f_x(a,b) + (y-b) f_y(a,b) \\
&+ \frac{1}{2}(x-a)^2 f_{xx}(a,b) + \frac{1}{2}(y-b)^2 f_{yy}(a,b) \\
&+ (x-a)(y-b) f_{xy}(a,b).
\end{aligned}
$$

- Gradient: $\nabla f(x,y) = \begin{bmatrix} \frac{\partial f}{\partial x}(x,y) \\ \frac{\partial f}{\partial y}(x,y) \end{bmatrix}$
- Hessian: $\mathbf{H}(x,y) = \begin{bmatrix} \frac{\partial^2 f}{\partial x^2}(x,y) & \frac{\partial^2 f}{\partial x \partial y}(x,y) \\ \frac{\partial^2 f}{\partial x \partial y}(x,y) & \frac{\partial^2 f}{\partial y^2}(x,y) \end{bmatrix}$
- Taylor series around $\mathbf{x}_0$ in general:

$$
f(\mathbf{x}) \approx f(\mathbf{x}_0) + (\mathbf{x}-\mathbf{x}_0)^T \nabla f(\mathbf{x}_0) + \frac{1}{2}(\mathbf{x}-\mathbf{x}_0)^T \mathbf{H}(\mathbf{x}_0)(\mathbf{x}-\mathbf{x}_0)
$$

Example

- Taylor series in 2-d around (a, b):

$$\begin{aligned} f(x,y) \approx f(a,b) &+ (x-a)f_x(a,b) + (y-b)f_y(a,b) \\ &+ \frac{1}{2}(x-a)^2 f_{xx}(a,b) + \frac{1}{2}(y-b)^2 f_{yy}(a,b) \\ &+ (x-a)(y-b)f_{xy}(a,b). \end{aligned}$$

- $f(x, y) = e^x e^y$ around $(0, 0)$

Example

- Taylor series in 2-d around (a, b):

$$\begin{aligned} f(x,y) \approx f(a,b) &+ (x-a)f_x(a,b) + (y-b)f_y(a,b) \\ &+ \frac{1}{2}(x-a)^2 f_{xx}(a,b) + \frac{1}{2}(y-b)^2 f_{yy}(a,b) \\ &+ (x-a)(y-b)f_{xy}(a,b). \end{aligned}$$

- $f(x, y) = e^x e^y$ around $(0, 0)$
- $f_x(0,0) = f_y(0,0) = f_{xx}(0,0) = f_{yy}(0,0) = f_{xy}(0,0) = 1$
- $f(x,y) \approx 1 + x + y + \frac{x^2}{2} + \frac{y^2}{2} + xy$
- $\nabla f(0,0) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$, $\mathbf{H}(0,0) = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$
- $f(x,y) \approx 1 + \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}$

Taylor series in 2-d

- $p_2(x,y) = c_{0,0} + c_{1,0}(x-a) + c_{0,1}(y-b) + c_{2,0}(x-a)^2 + c_{0,2}(y-b)^2 + c_{1,1}(x-a)(y-b)$

Taylor series in 2-d

- $p_2(x,y) = c_{0,0} + c_{1,0}(x-a) + c_{0,1}(y-b) + c_{2,0}(x-a)^2 + c_{0,2}(y-b)^2 + c_{1,1}(x-a)(y-b)$
- Match $f(x,y)$ at (a,b) for derivatives
- $f(a,b) = p_2(a,b) = c_{0,0}$
- $f_x(a,b) = \frac{\partial p_2(a,b)}{\partial x} = c_{1,0}$
- $f_y(a,b) = \frac{\partial p_2(a,b)}{\partial y} = c_{0,1}$
- $f_{xx}(a,b) = \frac{\partial^2 p_2(a,b)}{\partial x^2} = 2c_{2,0}$
- $f_{yy}(a,b) = \frac{\partial^2 p_2(a,b)}{\partial y^2} = 2c_{0,2}$
- $f_{xy}(a,b) = \frac{\partial^2 p_2(a,b)}{\partial x \partial y} = c_{1,1}$.

Nonlinear systems

- Linear systems:

$$
\begin{aligned}
x + y &= 10, \\
2x + 4y &= 26.
\end{aligned}
$$

- Nonlinear systems:

$$
\begin{aligned}
4x^2 + y^2 - 4 &= 0, \\
x + y - \sin(x - y) &= 0.
\end{aligned}
$$

- **Question**: Does grid searching work?

Nonlinear systems

- Linear systems:

$$\begin{aligned} x + y &= 10, \\ 2x + 4y &= 26. \end{aligned}$$

- Nonlinear systems:

$$\begin{aligned} 4x^2 + y^2 - 4 &= 0, \\ x + y - \sin(x - y) &= 0. \end{aligned}$$

- **Question**: Does grid searching work? Yes, with n^d function evaluation
- **Question**: Bisection?

Nonlinear systems

- Linear systems:

$$
\begin{aligned}
x + y &= 10, \\
2x + 4y &= 26.
\end{aligned}
$$

- Nonlinear systems:

$$
\begin{aligned}
4x^2 + y^2 - 4 &= 0, \\
x + y - \sin(x - y) &= 0.
\end{aligned}
$$

- **Question**: Does grid searching work? Yes, with n^d function evaluation
- **Question**: Bisection? No.
- **Question**: Newton?

Newton in 2-d

- Solve

$$f_1(x,y) = 0,$$
$$f_2(x,y) = 0.$$

- With initial guess (a,b)

$$f_1(x,y) \approx f_1(a,b) + (x-a)\frac{\partial f_1}{\partial x}(a,b) + (y-b)\frac{\partial f_1}{\partial y}(a,b),$$
$$f_2(x,y) \approx f_2(a,b) + (x-a)\frac{\partial f_2}{\partial x}(a,b) + (y-b)\frac{\partial f_2}{\partial y}(a,b).$$

- Suppose (x^*, y^*) is the root

$$0 \approx f_1(a,b) + x^*\frac{\partial f_1}{\partial x}(a,b) + y^*\frac{\partial f_1}{\partial y}(a,b) - a\frac{\partial f_1}{\partial x}(a,b) - b\frac{\partial f_1}{\partial y}(a,b),$$
$$0 \approx f_2(a,b) + x^*\frac{\partial f_2}{\partial x}(a,b) + y^*\frac{\partial f_2}{\partial y}(a,b) - a\frac{\partial f_2}{\partial x}(a,b) - b\frac{\partial f_2}{\partial y}(a,b)$$

Newton in 2-d - continued

- Solve

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} = \begin{bmatrix} f_1(a,b) \\ f_2(a,b) \end{bmatrix} + \begin{bmatrix} \frac{\partial f_1}{\partial x}(a,b) & \frac{\partial f_1}{\partial y}(a,b) \\ \frac{\partial f_2}{\partial x}(a,b) & \frac{\partial f_2}{\partial y}(a,b) \end{bmatrix} \begin{bmatrix} x^* - a \\ y^* - b \end{bmatrix}$$

- $\begin{bmatrix} x^* \\ y^* \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix} - \begin{bmatrix} \frac{\partial f_1}{\partial x}(a,b) & \frac{\partial f_1}{\partial y}(a,b) \\ \frac{\partial f_2}{\partial x}(a,b) & \frac{\partial f_2}{\partial y}(a,b) \end{bmatrix}^{-1} \begin{bmatrix} f_1(a,b) \\ f_2(a,b) \end{bmatrix}$
- Jacobian matrix: $\mathbf{J}(a,b) = \begin{bmatrix} \frac{\partial f_1}{\partial x}(a,b) & \frac{\partial f_1}{\partial y}(a,b) \\ \frac{\partial f_2}{\partial x}(a,b) & \frac{\partial f_2}{\partial y}(a,b) \end{bmatrix}$
- Newton's iteration:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix} - \mathbf{J}^{-1}(x_n, y_n) \begin{bmatrix} f_1(x_n, y_n) \\ f_2(x_n, y_n) \end{bmatrix}.$$

Example

- Nonlinear systems:

$$
\begin{aligned}
f_1(x, y) &= 4x^2 + y^2 - 4 = 0, \\
f_2(x, y) &= x - \sin(x - y) = 0.
\end{aligned}
$$

- Jacobian matrix:

Example

- Nonlinear systems:

$$f_1(x,y) = 4x^2 + y^2 - 4 = 0,$$
$$f_2(x,y) = x - \sin(x-y) = 0.$$

- Jacobian matrix:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial f_1}{\partial x}(x,y) & \frac{\partial f_1}{\partial y}(x,y) \\ \frac{\partial f_2}{\partial x}(x,y) & \frac{\partial f_2}{\partial y}(x,y) \end{bmatrix}$$
$$= \begin{bmatrix} 8x & 2y \\ 1-\cos(x-y) & \cos(x-y) \end{bmatrix}.$$

- Newton's iteration:

$$\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} = \begin{bmatrix} x_n \\ y_n \end{bmatrix}$$
$$- \begin{bmatrix} 8x_n & 2y_n \\ 1-\cos(x_n-y_n) & \cos(x_n-y_n) \end{bmatrix}^{-1} \begin{bmatrix} 4x_n^2 + y_n^2 - 4 \\ x_n - \sin(x_n-y_n) \end{bmatrix}$$

Newton - general cases

- $f_i(\mathbf{x}) \approx f_i(\mathbf{x}_0) + \nabla f_i^T(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$
- $\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{J}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$
- $\mathbf{0} = \mathbf{f}(\mathbf{x}^*) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{J}(\mathbf{x}_0)(\mathbf{x}^* - \mathbf{x}_0)$
- $\mathbf{x}^* \approx \mathbf{x}_0 - \mathbf{J}^{-1}(\mathbf{x}_0)\mathbf{f}(\mathbf{x}_0)$
- $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{J}^{-1}(\mathbf{x}_n)\mathbf{f}(\mathbf{x}_n)$
- **Question**: Most expensive cost?

Newton - general cases

- $f_i(\mathbf{x}) \approx f_i(\mathbf{x}_0) + \nabla f_i^T(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$
- $\mathbf{f}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{J}(\mathbf{x}_0)(\mathbf{x} - \mathbf{x}_0)$
- $\mathbf{0} = \mathbf{f}(\mathbf{x}^*) \approx \mathbf{f}(\mathbf{x}_0) + \mathbf{J}(\mathbf{x}_0)(\mathbf{x}^* - \mathbf{x}_0)$
- $\mathbf{x}^* \approx \mathbf{x}_0 - \mathbf{J}^{-1}(\mathbf{x}_0)\mathbf{f}(\mathbf{x}_0)$
- $\mathbf{x}_{n+1} = \mathbf{x}_n - \mathbf{J}^{-1}(\mathbf{x}_n)\mathbf{f}(\mathbf{x}_n)$
- **Question**: Most expensive cost?
- Solve the linear system
- E.g. N-body simulation

$$m_i \frac{d^2\mathbf{x}_i}{dt^2} = -\sum_{j \neq i} \frac{G m_i m_j (\mathbf{x}_i - \mathbf{x}_j)}{\|\mathbf{x}_i - \mathbf{x}_j\|^3}.$$

- **Question**: Is it possible to avoid this calculation?

Fixed-point iteration

- $\mathbf{x}_{n+1} = \mathbf{g}(\mathbf{x}_n)$
- For simplicity, let's focus on the linear system only
- $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^n$ with n large
- **Question**: Can we avoid $\mathbf{A}^{-1}$?

Fixed-point iteration

- $\mathbf{x}_{n+1} = \mathbf{g}(\mathbf{x}_n)$
- For simplicity, let's focus on the linear system only
- $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^n$ with n large
- **Question**: Can we avoid $\mathbf{A}^{-1}$?
- $\mathbf{x} = \mathbf{b} + \mathbf{A}\mathbf{b} + \mathbf{A}^2\mathbf{b} + \dots$
- **Question**: How to implement it?

Fixed-point iteration

- $\mathbf{x}_{n+1} = \mathbf{g}(\mathbf{x}_n)$
- For simplicity, let's focus on the linear system only
- $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^n$ with n large
- **Question**: Can we avoid $\mathbf{A}^{-1}$?
- $\mathbf{x} = \mathbf{b} + \mathbf{A}\mathbf{b} + \mathbf{A}^2\mathbf{b} + \dots$
- **Question**: How to implement it?
- $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{b}$
- Error: $\mathbf{e}_{n+1} = \mathbf{A}\mathbf{e}_n$
- **Question**: How fast does the error decay?

Fixed-point iteration

- $\mathbf{x}_{n+1} = \mathbf{g}(\mathbf{x}_n)$
- For simplicity, let's focus on the linear system only
- $(\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{b}$, $\mathbf{A} \in \mathbb{R}^n$ with n large
- **Question**: Can we avoid $\mathbf{A}^{-1}$?
- $\mathbf{x} = \mathbf{b} + \mathbf{A}\mathbf{b} + \mathbf{A}^2\mathbf{b} + \dots$
- **Question**: How to implement it?
- $\mathbf{x}_{n+1} = \mathbf{A}\mathbf{x}_n + \mathbf{b}$
- Error: $\mathbf{e}_{n+1} = \mathbf{A}\mathbf{e}_n$
- **Question**: How fast does the error decay?
- Suppose $\mathbf{A} = \mathbf{V}\mathbf{D}\mathbf{V}^{-1}$
- $\mathbf{e}_n = \mathbf{V}\mathbf{D}^n\mathbf{V}^{-1}$
- Focus on the eigenvalue with the largest magnitude

Convergence analysis

- Taylor series in 2d:

$$f(x,y) - f(x_0,y_0) \approx \frac{\partial f}{\partial x}(x_0,y_0)(x-x_0) + \frac{\partial f}{\partial y}(x_0,y_0)(y-y_0).$$

- Matrix equation for vector functions:

$$\begin{bmatrix} f_1(x,y) - f_1(x_0,y_0) \\ f_2(x,y) - f_2(x_0,y_0) \end{bmatrix} \approx \begin{bmatrix} \frac{\partial f_1(x_0,y_0)}{\partial x} & \frac{\partial f_1(x_0,y_0)}{\partial y} \\ \frac{\partial f_2(x_0,y_0)}{\partial x} & \frac{\partial f_2(x_0,y_0)}{\partial y} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix}.$$

- Inf-norm of a vector $\|\mathbf{x}\|_\infty = \max_i |x_i|$
- Want: $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)\|_\infty \le \lambda \|\mathbf{x} - \mathbf{x}_0\|_\infty$

Convergence analysis

- $|f_1(x,y)-f_1(x_0,y_0)| \leq \left|\frac{\partial f_1(x_0,y_0)}{\partial x}\right| |x-x_0| + \left|\frac{\partial f_1(x_0,y_0)}{\partial y}\right| |y-y_0|$
- $|f_2(x,y)-f_2(x_0,y_0)| \leq \left|\frac{\partial f_2(x_0,y_0)}{\partial x}\right| |x-x_0| + \left|\frac{\partial f_2(x_0,y_0)}{\partial y}\right| |y-y_0|$
- $|f_i(x,y) - f_i(x_0,y_0)| \leq \left(\left|\frac{\partial f_i(x_0,y_0)}{\partial x}\right| + \left|\frac{\partial f_i(x_0,y_0)}{\partial y}\right|\right) \max(|x-x_0|, |y-y_0|)$
- $\max_i |f_i(x,y) - f_i(x_0,y_0)| \leq \max_i \left(\left|\frac{\partial f_i(x_0,y_0)}{\partial x}\right| + \left|\frac{\partial f_i(x_0,y_0)}{\partial y}\right|\right) \max(|x-x_0|, |y-y_0|)$
- Inf-norm of a matrix $\|\mathbf{A}\|_\infty = \max_i \sum_{j=1}^n |A_{i,j}|$
- $\|\mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}_0)\|_\infty \leq \|\mathbf{J}(\mathbf{x}_0)\|_\infty \|\mathbf{x} - \mathbf{x}_0\|_\infty$
- Can be easily generalized to high-dimensions

Convergence analysis

- Assumptions:
  - Suppose $\mathbf{x}^*$ is the solution to $\mathbf{x} = \mathbf{g}(\mathbf{x})$
  - $\mathbf{g}(\mathbf{x})$ is continuously differentiable in some neighborhood around $\mathbf{x}^*$
  - $\lambda = \|\mathbf{J}(\mathbf{x}^*)\|_\infty \leq 1$
- **Question**: How to analyze the convergence?

Convergence analysis

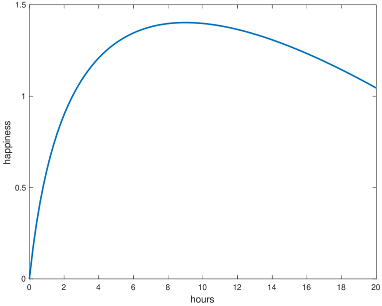
- Assumptions:
  - Suppose $\mathbf{x}^*$ is the solution to $\mathbf{x} = \mathbf{g}(\mathbf{x})$
  - $\mathbf{g}(\mathbf{x})$ is continuously differentiable in some neighborhood around $\mathbf{x}^*$
  - $\lambda = \|\mathbf{J}(\mathbf{x}^*)\|_\infty \leq 1$
- **Question**: How to analyze the convergence?
- $\mathbf{x}^* - \mathbf{x}_{n+1} = \mathbf{g}(\mathbf{x}^*) - \mathbf{g}(\mathbf{x}_n)$
- $\|\mathbf{g}(\mathbf{x}) - \mathbf{g}(\mathbf{x}_0)\|_\infty \leq \|\mathbf{J}(\mathbf{x}_0)\|_\infty \|\mathbf{x} - \mathbf{x}_0\|_\infty$
- $\|\mathbf{x}^* - \mathbf{x}_{n+1}\|_\infty \leq \lambda \|\mathbf{x}^* - \mathbf{x}_n\|_\infty \leq \lambda^{n+1} \|\mathbf{x}^* - \mathbf{x}_0\|_\infty$

Optimization problem

- In real-life, many problems are formulated to optimization problem

$$\min_x f(x)$$

- E.g. happiness vs hours



- **Question**: How to solve this numerically?

Newton's method for optimization

- One condition of the minimizer: $f'(x^*) = 0$
- Reformulate it to a root-finding problem
- $x_{n+1} = x_n - \frac{f'(x_n)}{f''(x_n)}$
- Newton's method is not the most popular method in optimization
- Other methods include gradient descent
- Will cover in optimization courses