CS450: Numerical Analysis Nonlinear Equation Systems

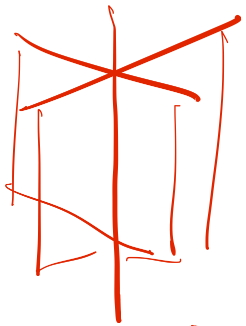
Howard Hao Yang

Assistant Professor, ZJU-UIUC Institute

29/11/2023

Recap: Linear Equation Systems

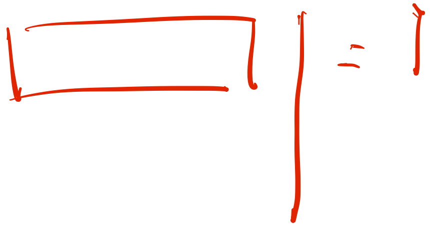
- Given an $m \times n$ matrix $\boldsymbol{A}$ and an m-dimensional vector $\boldsymbol{b}$, we aim to find n-dimensional vector $\boldsymbol{x}$ such that $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$
  - Linear combination of columns of $\boldsymbol{A}$ to yield $\boldsymbol{b}$
- Geometric interpretation: Line/plane intersection



- How many solutions might it have?
  - Depends on $\boldsymbol{A}$
  - But always three possible cases: zero, one, or infinity

$m = n$. ~~Gt~~ nonsingular matrix A

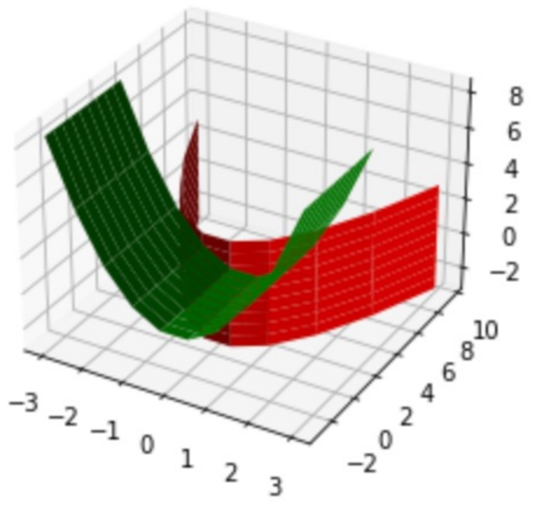
$n > m$.



Nonlinear Equation Systems

$f_1(x_1, \dots, x_n) = 0$
$f_2(x_1, \dots, x_n) = 0$
$\vdots$
$f_n(x_1, \dots, x_n) = 0$

- Given a mapping $\boldsymbol{f}: R^n \to R^n$, we aim to find n-dimensional vector $\boldsymbol{x}$ such that $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$
  - If we look for solution to $\tilde{\boldsymbol{f}}(\boldsymbol{x}) = \boldsymbol{y}$, simply consider $\boldsymbol{f}(\boldsymbol{x}) = \tilde{\boldsymbol{f}}(\boldsymbol{x}) - \boldsymbol{y}$
- Geometric interpretation: Curve intersection
- How many solutions might it have?
  - Depends on the equations
  - Can be any possible integers



Illustrative Example

- Consider the system of equations in two dimensions

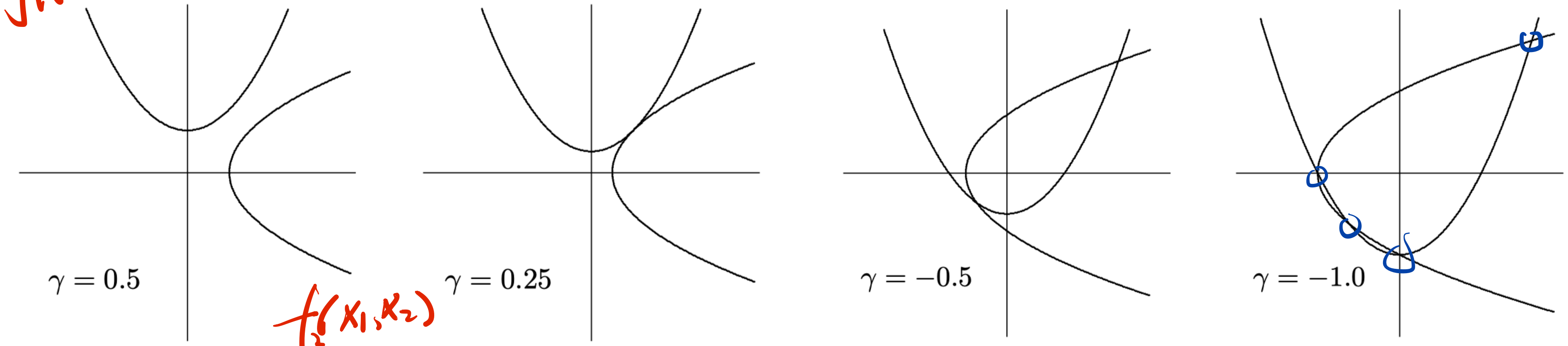
$$\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} x_1^2 - x_2 + \gamma \\ -x_1 + x_2^2 + \gamma \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$f_1(x_1, x_2) = x_1^2 - x_2 + \gamma = 0$

$f_2(x_1, x_2) = -x_1 + x_2^2 + \gamma = 0$

- The number of solutions depends on the particular value of γ

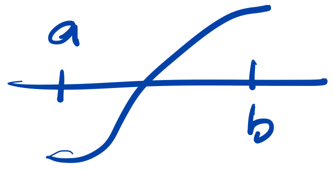
$f_1(x_1, x_2)$



$f_2(x_1, x_2)$

Existence (1)

- Given $\boldsymbol{f}$, when would there exist a variable $\boldsymbol{x}$ that solves $\boldsymbol{f}(\boldsymbol{x}) = \mathbf{0}$?
- 1-Dimension case
  - (**Intermediate Value Theorem**) if $f: R \to R$ is continuous on $[a, b]$, and c lies between $f(a)$ and $f(b)$, then there is a value $x^* \in [a, b]$, such that $f(x^*) = c$
  - Q: Is continuity necessary? Is closed interval necessary?
  - How can we use this to determine the existence of a root?
  - Can we determine the number of roots?
  - In general, how to find such an interval?



Existence (2)

$$\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

- Given $\boldsymbol{f}$, when would there exist a variable $\boldsymbol{x}$ that solves $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$?
- n-Dimension case

  $f_1, \dots, f_n$

  - (**Inverse Function Theorem**) if $\boldsymbol{f}: R^n \to R^n$ is continuously differentiable, if the Jacobian matrix $\{\boldsymbol{J}_f(\boldsymbol{x})\}_{ij} = \boldsymbol{\partial f}_i(\boldsymbol{x})/\boldsymbol{\partial x}_j$ is nonsingular at a point $\boldsymbol{x}^*$. Then, there is a neighborhood of $\boldsymbol{f}(\boldsymbol{x}^*)$ in which $\boldsymbol{f}^{-1}$ exists. That is, $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{y}$ has a solution for any $\boldsymbol{y}$ in that neighborhood of $\boldsymbol{f}(\boldsymbol{x}^*)$.

    $\Rightarrow x = f^{-1}(y)$

  - Issue: only local, not global. Besides, relies on calculating the Jacobian matrix.

Existence (3)

- Given $\boldsymbol{f}$, when would there exist a variable $\boldsymbol{x}$ that solves $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$?
- n-Dimension case

  $x = (x_1, \ldots, x_n)^T$
  $y = (y_1, \ldots, y_n)^T$

  - (**Contraction Mapping Theorem**) a function $\boldsymbol{g}: R^n \to R^n$ is called contractive if there exists $0 < \gamma < 1$ such that $\|\boldsymbol{g}(\boldsymbol{x}) - \boldsymbol{g}(\boldsymbol{y})\| \le \gamma \|\boldsymbol{x} - \boldsymbol{y}\|$. A fixed point of $\boldsymbol{g}$ is a point where $\boldsymbol{g}(\boldsymbol{x}) = \boldsymbol{x}$

    vector norm.

  - On a closed set $S \subset R^n$ with $\boldsymbol{g}(S) \subset S$ there exists a unique fixed point (why?)
  - Example: real-world map

Solving Nonlinear Equation (1)

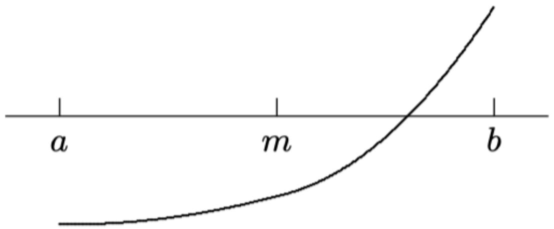
- Given a continuous function $f: R \to R$, how to find x such that $f(x) = 0$?
- Interval Bisection:

  1) start with an interval $[a, b]$ in which f changes sign

  2) cut off half of the interval, till reaches a certain level

Algorithm 5.1 Interval Bisection

```
while ((b − a) > tol) do
    m = a + (b − a)/2
    if sign(f(a)) = sign(f(m)) then
        a = m
    else
        b = m
    end
end
```



Solving Nonlinear Equation (2)

- Given a continuous function $f: R \to R$, how to find x such that $f(x) = 0$?
- Fixed Point Iteration:
  1) transform the equation into $g(x) = x$
  2) start with an initial guess x_0
  3) keep iterating $x_{k+1} = g(x_k)$
- Convergent criteria: if $x^* = g(x^*)$ is a fixed point and $|g(x^*)| < 1$, then the iterative scheme converges

$f(x)=0$

$\downarrow$

$\underbrace{x+f(x)}=x$

$g(x)=x$

Two Questions with this Iteration

- In general, for iterative algorithms, there are always two questions associated with the process
- Q1: when shall one stop?
  - We need a stopping criteria
- Q2: how fast the iteration converges?
  - We need to quantify the "speed of training"

Stopping Criteria

- Possible candidates for stopping the iteration
  - Run until $|f(x)| < \epsilon$, i.e., the residual is small
  - Run until $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| < \epsilon$
  - Run until $\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\| / \|\boldsymbol{x}_k\| < \epsilon$
- None of them is bulletproof, depends on the application
  - Run until iteration number exceeds a predetermined value is also a good strategy

Rate of Convergence

est. sol.

- Consider an iterative algorithm to solve $\boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{0}$, where $\boldsymbol{e}_k = \widehat{\boldsymbol{x}}_k - \boldsymbol{x}^*$ is the error in the k-th iteration. Assume $\boldsymbol{e}_k \to 0$ as $k \to \infty$, an iterative algorithm converges with rate r if

$$\frac{\|e_{k+1}\|}{\|e_k\|} = C$$

$$\lim_{k\to\infty} \frac{\|\boldsymbol{e}_{k+1}\|}{\|\boldsymbol{e}_k\|^r} = C \begin{cases} > 0 \\ < \infty \end{cases}$$

  - If $r = 1$, it is called *linear convergence* (Example: Power Method)
  - If $r > 1$, it is called *superlinear convergence*

    $1 < r < 2$
  - If $r = 2$, it is called *quadratic convergence* (Example: Rayleigh Quotient Iteration)

Example (0)

- If the errors at successive iterations of an iterative method are as follows, how would you characterize the convergence rate?

$a)$ $10^{-2}, 10^{-4}, 10^{-8}, 10^{-16}, \ldots$ $r = ?$ 2

$b)$ $10^{-2}, 10^{-4}, 10^{-6}, 10^{-8}, \ldots$ $r = ?$ 1

a).

$e_1 = 10^{-2}$

$e_2 = 10^{-4}$

$e_3 = 10^{-8}$

$\vdots$

$\frac{|e_2|}{|e_1|^2} = 1$

$\frac{|e_3|}{|e_2|^2} = 1$

$r = 2$

b)

$e_1 = 10^{-2}$

$e_2 = 10^{-4}$

$e_3 = 10^{-6}$

$\vdots$

$\frac{\|e_2\|}{\|e_1\|} = 10^{-2}$

$\frac{\|e_3\|}{\|e_2\|} = 10^{-2}$

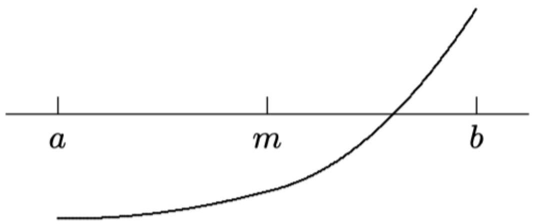
$r = 1$

Example (1)

- What is the convergence rate of Bisection Interval?

Algorithm 5.1 Interval Bisection

```
while ((b − a) > tol) do
    m = a + (b − a)/2
    if sign(f(a)) = sign(f(m)) then
        a = m
    else
        b = m
    end
end
```



$$e_1 \le \frac{b-a}{2}$$

$$e_2 \le \frac{b-a}{4}$$

$$\vdots$$

$$e_n \le \frac{b-a}{2^n}$$

- Linear with constant $\frac{1}{2}$ (can you see this?)

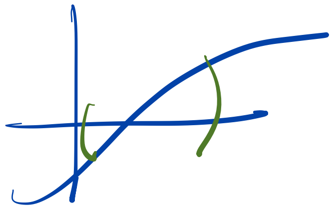
$r=1$

C $\qquad \dfrac{e_{k+1}}{e_k} \le \dfrac{1}{2}$

Characterizing Fixed-Point Iterations

$f(x)=0$

- For the fixed-point iteration $x_{k+1} = g(x_k)$, how to characterize the convergence rate?
- A simple (though not the most general) approach



  - Consider the derivative of g at the solution x^* (of $x = g(x)$)
  - If $x^* = g(x^*)$ and $|g'(x^*)| < 1$, then the fixed-point iteration is locally convergent with the constant $C = |g'(x^*)|$

$$|x_{k+1} - x^*| = |g(x_k) - g(x^*)|$$

$$\leq \underbrace{|g'(x^*)|}_{C} \cdot |x_k - x^*|$$

Example (2)

- Consider the following nonlinear equation

$$f(x) = x^2 - x - 2 = 0$$

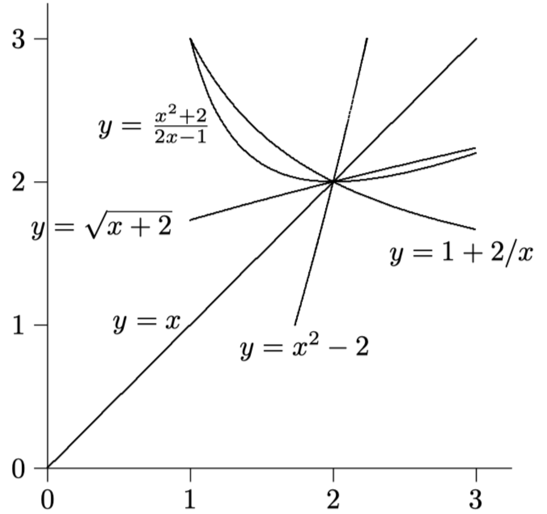
- We want to transform this equation into the form $g(x) = x$
- Equivalent fixed-point problems include

1) $x^2 - 2 = x \quad \Rightarrow \quad g(x) = x^2 - 2$

2) $x^2 = x + 2 \quad \Rightarrow \quad x = \sqrt{x+2} \quad \Rightarrow \quad g(x) = \sqrt{x+2}$

3) $x^2 = x + 2 \quad \Rightarrow \quad x = 1 + \frac{2}{x} \quad \Rightarrow \quad g(x) = 1 + \frac{2}{x}$

4) $x^2 + 2 = 2x^2 - x \quad \Rightarrow \quad x = \frac{x^2+2}{2x-1} \quad \Rightarrow \quad g(x) = \frac{x^2+2}{2x-1}$



Exercise

- Consider the following nonlinear equation

$$f(x) = x^2 - x - 2 = 0$$

$x^* = 2$

- Transform this equation into the fixed-point form $g(x) = x$

$C = |f'(x^*)|$

- Find constant C of convergence for the following transformations

1) $x^2 - 2 = x \Rightarrow g(x) = x^2 - 2$ → $C = f'(x)\big|_{x^*=2} = 2x^* = 4$

② 2) $x^2 = x + 2 \Rightarrow x = \sqrt{x+2} \Rightarrow g(x) = \sqrt{x+2}$ → $C = f'(x) = \frac{1}{2\sqrt{x+2}}\Big|_{x^*=2} = \frac{1}{4}$

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

③ 3) $x^2 = x + 2 \Rightarrow x = 1 + \frac{2}{x} \Rightarrow g(x) = 1 + \frac{2}{x}$ → $C = f'(x) = -\frac{2}{x^2}\Big|_{x^*=2} = -\frac{1}{2}$

① 4) $x^2 + 2 = 2x^2 - x \Rightarrow x = \frac{x^2+2}{2x-1} \Rightarrow g(x) = \frac{x^2+2}{2x-1}$ → $C = f'(x)$

$= \frac{2x}{2x-1} - \frac{2(x^2+2)}{(2x-1)^2}\Big|_{x^*=2}$

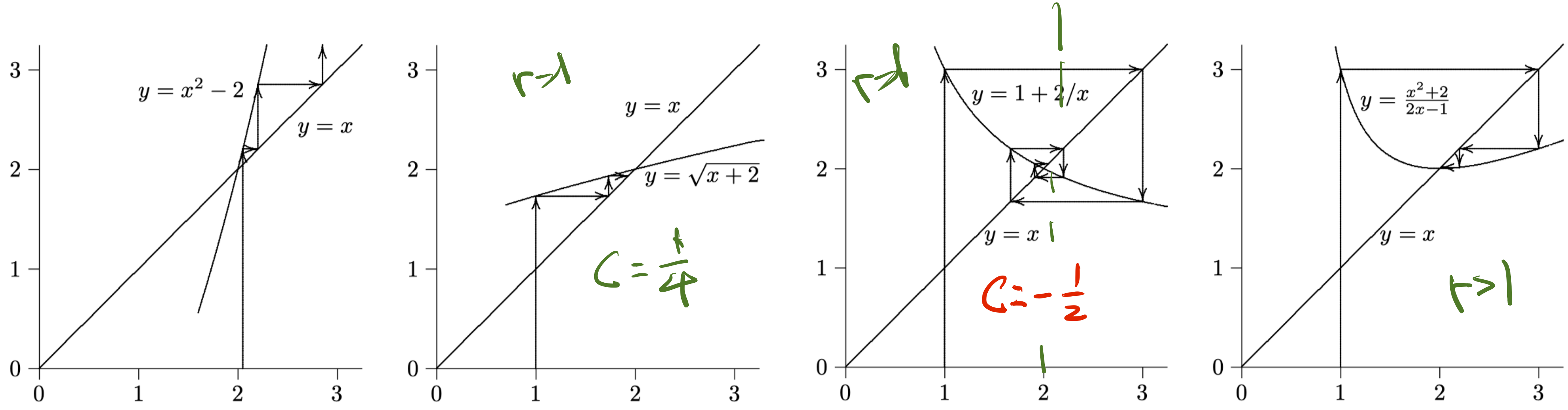
$= 0$

Illustrative Example

- Consider the following nonlinear equation

$$f(x) = x^2 - x - 2 = 0$$

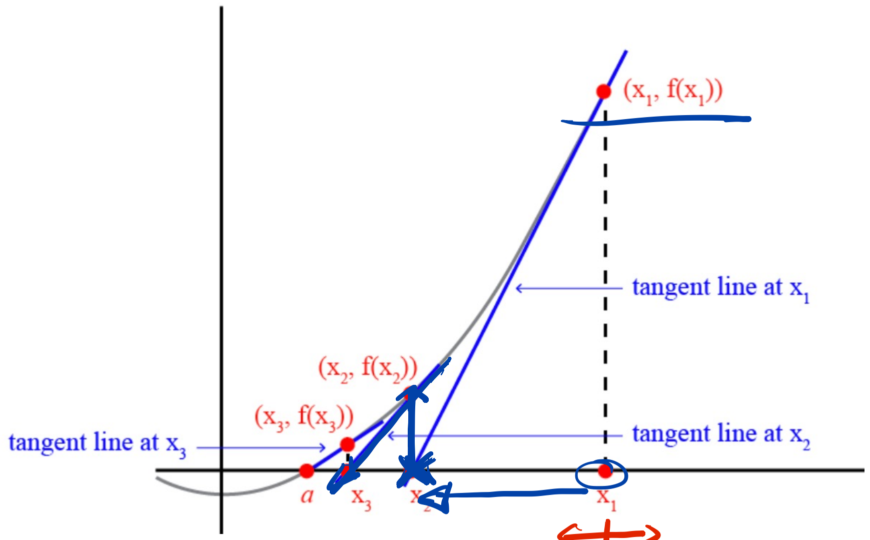
- Transform this equation into the form $g(x) = x$, and solve by fixed-point iteration—different forms converges (or diverges) differently



Newton's Method

$f(x)=0$

$\vec{f}(x)=0$



$x_k+\Delta_k$

- Idea:

    1) suppose at every iteration (say the k-th), the next estimation is updated by the current one with a small difference $x_{k+1} = x_k + \Delta_k$

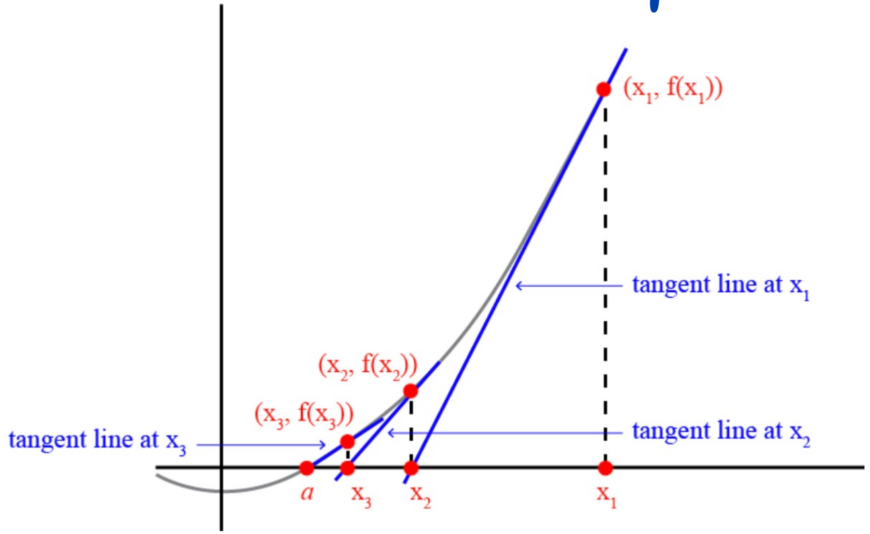
    2) approximate f by a linear function, i.e.,

$$f(x_{k+1}) = f(x_k + \Delta_k) \approx f(x_k) + f'(x_k)\Delta_k = 0$$

$$\Delta_k = -\frac{f(x_k)}{f'(x_k)}$$

Newton's Method (Cont'd)

$f(x)=0$



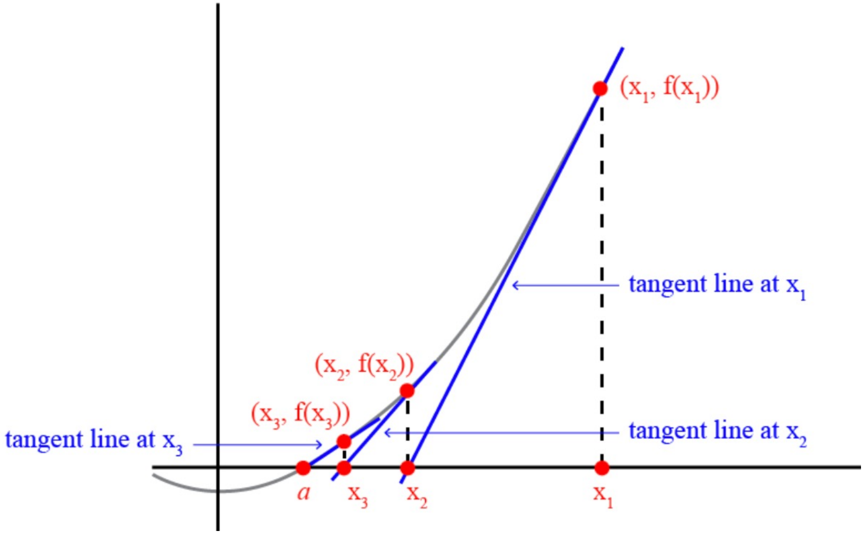
- Idea:

3) find root of this linear function in terms of Δ_k: $f(x_k) + f'(x_k)\Delta_k = 0 \rightarrow \Delta_k = -\frac{f(x_k)}{f'(x_k)}$

4) update the evaluation toward the zero-intersecting point

$$x_{k+1} = x_k + \Delta_k = x_k - \frac{f(x_k)}{f'(x_k)}$$

Newton's Method (Cont'd)



- Algorithm details:

  1) start with an initial guess x_0

  2) keep iterating $x_{k+1} = x_k - \frac{f(x_k)}{f\prime(x_k)}$

Newton's Method: Convergence Rate

$f(x)=0$

$g(x) = x - \frac{f(x)}{f'(x)}$

$0 \quad \cancel{x^*} = g(x^*) = \cancel{x^*} - \frac{f(x^*)}{f'(x^*)}$

$\rightarrow f(x^*)=0$

$C = g'(x^*) = 1 - \frac{f'(x^*)}{f'(x^*)} - \frac{f(x^*)^{=0}}{f''(x^*)} = 0$

- Algorithm details:
  1) start with an initial guess x_0
  2) keep iterating $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$
- We can form $g(x) = x - \frac{f(x)}{f'(x)}$, then the update in Newton's method again constitutes a fixed-point iteration, as $x_{k+1} = g(x_k)$
- What would be the constant $C = |g'(x^*)|$ where $x^* = g(x^*)$?

  $r > 1$
- How to characterize the convergence rate?

Newton's Method: Convergence Rate

$f(x)=0$

- Let ① $g(x) = x - \frac{f(x)}{f'(x)}$, $x^* = g(x^*)$, and ② $x_{k+1} = g(x_k)$
  - The error bound over each iteration is

$$e_{k+1} = x_{k+1} - x^* = g(x_k) - \cancel{g(x^*)}$$

  - Expand $g(x_k)$ around x^*, we have

$$g(x_k) = \cancel{g(x^*)} + \underbrace{\cancel{g'(x^*)|x_k - x^*|}}_{=0} + \frac{g''(\xi)}{2}|x_k - x^*|^2, \qquad \xi \in [x^*, x_k]$$

  - Then, we have the convergence rate as follows (which is quadratic)

$$e_{k+1} = g(x_k) - g(x^*) = \frac{g''(\xi)}{2}|x_k - x^*|^2 = C \cdot e_k^2$$

$\Rightarrow r=2$

Illustrative Example

- How to quickly evaluate $\sqrt[7]{1000}$? – find the root of $f(x) = x^7 - 1000 = 0$
- Because $f'(x) = 7x^6$, the algorithm goes as

$r=2 \to r=3$

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = x_k - \frac{x_k^7 - 1000}{7x_k^6}$$

- Initialize with a random guess $x_1 = 3$

$$x_2 = x_1 - \frac{x_1^7 - 1000}{7x_1^6} = (3) - \frac{(3)^7 - 1000}{7(3)^6} = 2.76739173$$

$$x_3 = x_2 - \frac{x_2^7 - 1000}{7x_2^6} = (2.76739173) - \frac{(2.76739173)^7 - 1000}{7(2.76739173)^6} = 2.69008741$$

$$x_4 = x_3 - \frac{x_3^7 - 1000}{7x_3^6} = (2.69008741) - \frac{(2.69008741)^7 - 1000}{7(2.69008741)^6} = 2.68275645$$

$$x_5 = x_4 - \frac{x_4^7 - 1000}{7x_4^6} = (2.68275645) - \frac{(2.68275645)^7 - 1000}{7(2.68275645)^6} = 2.68269580$$

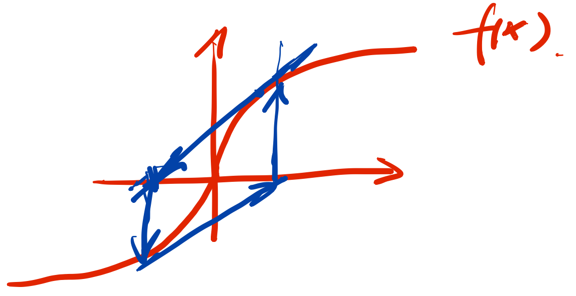
$$x_6 = x_5 - \frac{x_5^7 - 1000}{7x_5^6} = (2.68269580) - \frac{(2.68269580)^7 - 1000}{7(2.68269580)^6} = 2.68269580$$

Newton’s Method’s Issues (1)

- Newton’s method might be trapped in never ending cycles
- Example: consider finding the root of the following function, if we use Newton’s method with any initial guess $x_0 \neq 0$, we are trapped in a loop

① $f'(x) = \begin{cases} \frac{1}{2\sqrt{x}}, & x > 0 \\ \frac{1}{2\sqrt{-x}}, & x < 0 \end{cases}$

$$f(x) = \begin{cases} \sqrt{x}, & x \geq 0 \\ -\sqrt{-x}, & x < 0 \end{cases}$$



② $\frac{f(x)}{f'(x)} = \begin{cases} 2x, & x > 0 \\ 2x, & x < 0 \end{cases}$

③ $x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)} = -x_k \Rightarrow x_{k+2} = x_k$

Newton’s Method’s Issues (2)

ZJU-UIUC INSTITUTE

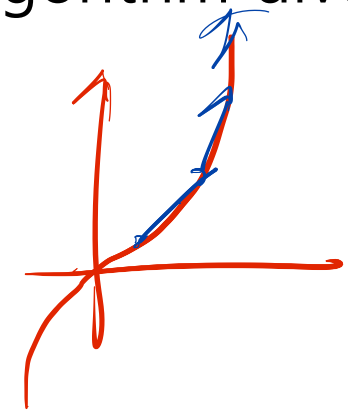
- Newton’s method might diverge
- Example: consider finding the root of the following function, if we use Newton’s method with any initial guess $x_0 \neq 0$, the algorithm diverges

$$f(x) = \sqrt[3]{x}$$

① $f'(x) = \frac{1}{3} \cdot x^{-\frac{2}{3}}$

② $\frac{f(x)}{f'(x)} = 3 \cdot x^{\frac{1}{3}}$

③ $x_{k+1} = x_k - \frac{f'(x_k)}{x_k} = \sqrt[3]{x_k} - 3\sqrt[3]{x_k} = -2\sqrt[3]{x_k}$



Newton's Method's Issues (3)

- Newton's method might not even kick off starting
- Example: consider finding the root of the following function, if we use Newton's method with an initial guess $x_0 = 0$, the algorithm cannot start

$$f(x) = x^3 - 1$$

$$f'(x_0) = 3x_0^2 = 0$$

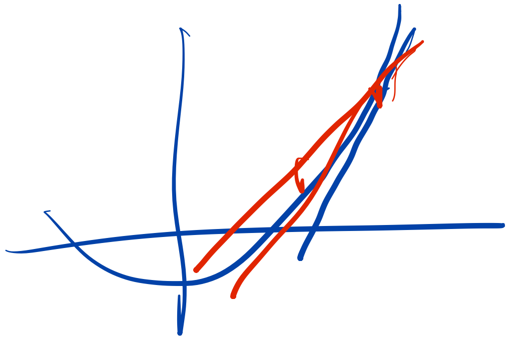
$$\frac{f(x)}{f'(x)}$$

Newton's Method: Convergence?

- **Theorem**: Suppose that $f(x) \in C^2[a, b]$. If there exists $x \in (a, b)$ such that $f(x) = 0$ and $f'(x) \neq 0$, then there exists $\delta > 0$ such that Newton's method generates a sequence $\{x_n\}_{n=1}^{\infty}$ converge to x for any initial approximation $x_0 \in [x - \delta,\ x + \delta]$.

Secant Method

- What would Newton's method without the use of derivative look like?
- Approximate $f'(x_k) \approx \frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}}$



- Algorithm details:
  1) start with an initial guess x_0
  2) keep iterating $x_{k+1} = g(x_k) = x_k - \frac{f(x_k)}{\frac{f(x_k)-f(x_{k-1})}{x_k-x_{k-1}}}$
- Secant (and similar methods) are called Quasi-Newton Methods

Secant Method (Cont'd)

- Algorithm details:
  1) start with an initial guess x_0
  2) keep iterating $x_{k+1} = g(x_k) = x_k - \frac{f(x_k)}{\frac{f(x_k)-f(x_{k-1})}{x_k - x_{k-1}}}$
- What is the convergence rate?
  - Rate of convergence is roughly $\frac{1+\sqrt{5}}{2} \approx 1.618$
- Setback?
  - Slower convergence rate
  - Need two starting points

Learning Objectives

- Nonlinear equation systems: Finding intersections of curves
- Existence: different criteria to judge
- Sensitivity: duality to the evaluation of a function
- Solving nonlinear equations: Bisection Interval, Fixed-Point Iteration, Newton's Method, Secant method
- Stopping criteria—no general bulletproof
- Convergence rate