# 11. Nonlinear equations with one variable

- definition and examples
- bisection method
- Newton's method
- secant method

# Definition and examples

$x$ is a *zero* (or *root*) of a function $f$ if $f(x) = 0$

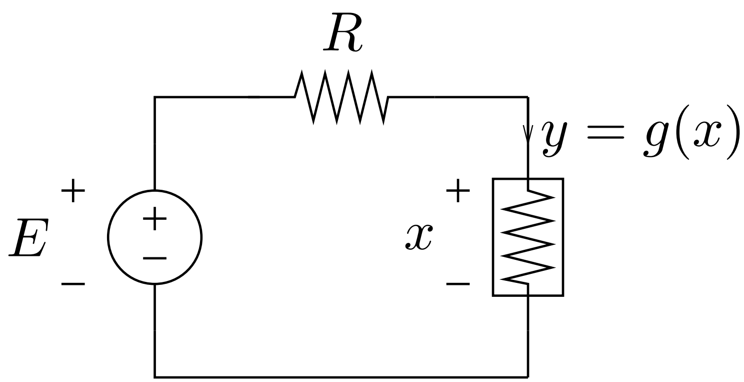
**examples**

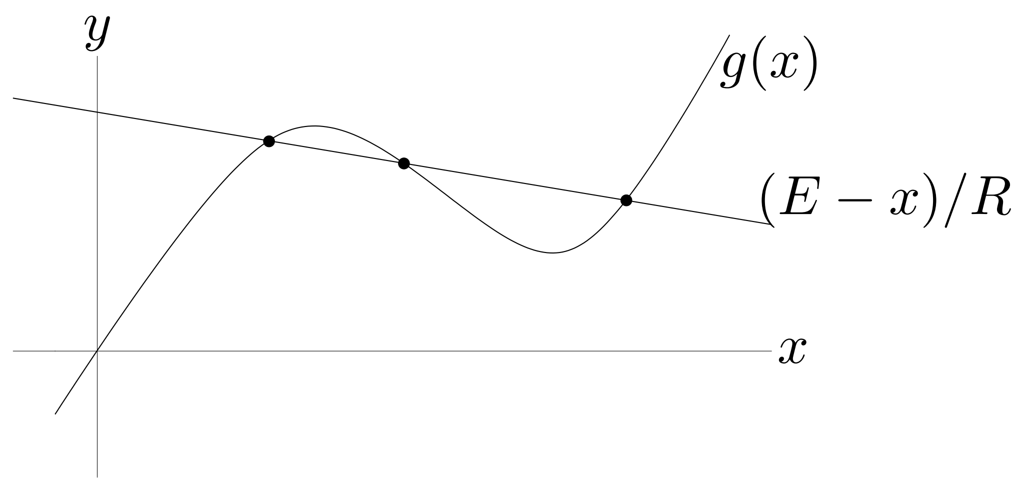
- $f(x) = e^x$ has no zeros
- $f(x) = e^x - e^{-x}$ has one zero
- $f(x) = e^x - e^{-x} - 3x$ has three zeros
- $f(x) = \cos x$ has infinitely many zeros

*cf.*, one linear equation in one variable $ax = b$

- a unique solution if $a \neq 0$
- no solution if $a = 0$, $b \neq 0$
- any $x \in \mathbf{R}$ is a solution if $a = b = 0$

# Example: nonlinear static circuit




operating point satisfies

$$g(x) - \frac{E - x}{R} = 0$$

- one nonlinear equation in one variable $x$
- three solutions

# Characteristics of algorithms for nonlinear equations

**how $f$ is described**

- user provides subroutine to compute $f(x)$ (and possibly $f'(x)$) at $x$
- called 'black box' or 'oracle' model for describing $f$
- evaluating $f$ and $f'$ can be expensive ($e.g.$, require a circuit simulation)

**limitations of algorithms**

- there exist no algorithms that are guaranteed to find all solutions
- most algorithms find at most one solution
- need prior information from the user: $e.g.$, an interval that contains a zero, or a point near a solution

## methods for solving nonlinear equations are iterative

- generate a sequence of points $x^{(k)}$, $k = 0, 1, 2, \ldots$ that converge to a solution; $x^{(k)}$ is called the $k$th *iterate*; $x^{(0)}$ is the *starting point*
- computing $x^{(k+1)}$ from $x^{(k)}$ is called one *iteration* of the algorithm
- each iteration typically requires one evaluation of $f$ (or $f$ and $f'$) at $x^{(k)}$
- algorithms need a stopping criterion, $e.g.$, terminate if

$$|f(x^{(k)})| \leq \text{specified tolerance}$$

- speed of the algorithm depends on:
  - the cost of evaluating $f(x)$ (and possibly, $f'(x)$)
  - the number of iterations

# Analyzing speed of convergence

suppose $x^{(k)} \to x^\star$ with $f(x^\star) = 0$; how fast does $x^{(k)}$ go to $x^\star$?

**error** after $k$ iterations:

- **absolute error**: $|x^{(k)} - x^\star|$
- **relative error**: $|x^{(k)} - x^\star|/|x^\star|$ (defined if $x^\star \neq 0$)
- **number of correct digits:**

$$\left\lfloor -\log_{10}\left(\frac{|x^{(k)} - x^\star|}{|x^\star|}\right)\right\rfloor$$

(defined if $x^\star \neq 0$ and $|x^{(k)} - x^\star|/|x^\star| \leq 1$)

**rates of convergence** of a sequence $x^{(k)}$ with limit $x^\star$

- linear convergence: there exists a $c \in (0,1)$ such that

$$|x^{(k+1)} - x^\star| \le c\,|x^{(k)} - x^\star| \quad \text{for sufficiently large } k$$

- R-linear convergence: there exists $c \in (0,1)$, $M > 0$ such that

$$|x^{(k)} - x^\star| \le M c^k \quad \text{for sufficiently large } k$$

- quadratic convergence: there exists a $c > 0$ s.t.

$$|x^{(k+1)} - x^\star| \le c\,|x^{(k)} - x^\star|^2 \quad \text{for sufficiently large } k$$

- superlinear convergence: there exists a sequence $c_k$ with $c_k \to 0$ s.t.

$$|x^{(k+1)} - x^\star| \le c_k\,|x^{(k)} - x^\star| \quad \text{for sufficiently large } k$$

**interpretation** (if $x^\star \neq 0$): let

$$r^{(k)} = -\log_{10}\left(\frac{|x^{(k)} - x^\star|}{|x^\star|}\right)$$

($i.e.$, $r^{(k)} \approx$ the number of correct digits at iteration $k$)

- linear convergence: we gain roughly $-\log_{10} c$ correct digits per step

$$r^{(k+1)} \geq r^{(k)} - \log_{10} c$$

- quadratic convergence: for $k$ sufficiently large, number of correct digits roughly doubles in one step

$$r^{(k+1)} \geq -\log(c|x^\star|) + 2r^{(k)}$$

- superlinear convergence: number of correct digits gained per step increases with $k$

$$r^{(k+1)} - r^{(k)} \to \infty$$

**examples** (with $x^\star = 1$)

- $x^{(k)} = 1 + 0.5^k$ converges linearly (with $c = 1/2$):

$$\frac{|x^{(k+1)} - 1|}{|x^{(k)} - 1|} = \frac{2^k}{2^{k+1}} = \frac{1}{2}$$

- $x^{(k)} = 1 + 0.5^{2^k}$ converges quadratically (with $c = 1$)

$$\frac{|x^{(k+1)} - 1|}{|x^{(k)} - 1|^2} = \frac{(2^{2^k})^2}{2^{2^{k+1}}} = 1$$

- $x^{(k)} = 1 + (1/(k+1))^k$ converges superlinearly

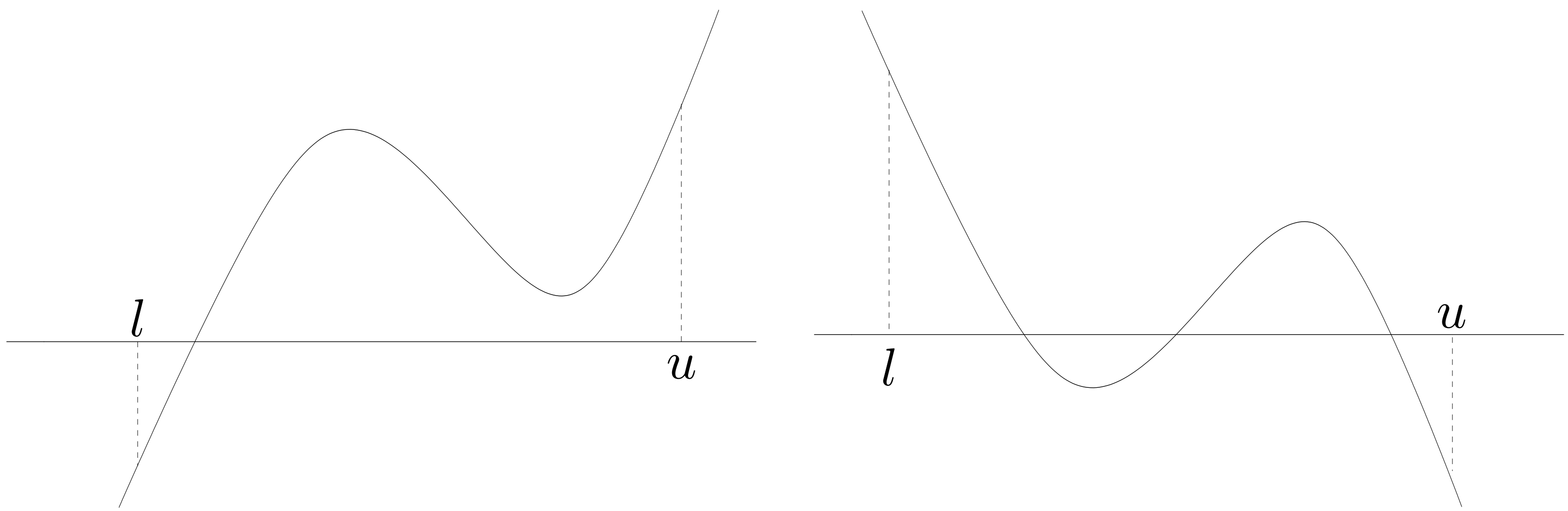
$$\frac{|x^{(k+1)} - 1|}{|x^{(k)} - 1|} = \frac{(k+1)^k}{(k+2)^{k+1}} \to 0$$

| $k$ | $1 + 0.5^k$ | $1 + 0.5^{2^k}$ | $1 + (1/(k+1)^k)$ |
|---|---|---|---|
| 0 | 2.00000000000000 | 1.50000000000000 | 2.00000000000000 |
| 1 | 1.50000000000000 | 1.25000000000000 | 1.50000000000000 |
| 2 | 1.25000000000000 | 1.06250000000000 | 1.11111111111111 |
| 3 | 1.12500000000000 | 1.00390625000000 | 1.01562500000000 |
| 4 | 1.06250000000000 | 1.00001525878906 | 1.00160000000000 |
| 5 | 1.03125000000000 | 1.00000000023283 | 1.00012860082305 |
| 6 | 1.01562500000000 | 1.00000000000000 | 1.00000849985975 |
| 7 | 1.00781250000000 | 1.00000000000000 | 1.00000047683716 |
| 8 | 1.00390625000000 | 1.00000000000000 | 1.00000002323057 |
| 9 | 1.00195313125000 | 1.00000000000000 | 1.00000000100000 |
| 10 | 1.00097656250000 | 1.00000000000000 | 1.00000000003855 |

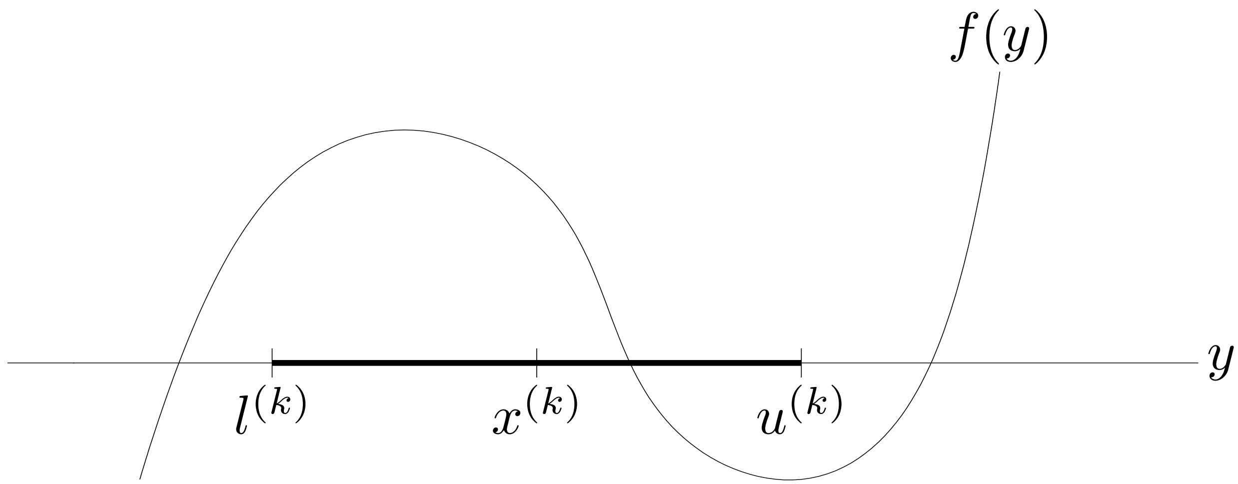
- sequence 1: we gain roughly $-\log_{10}(c) = 0.3$ correct digits per step
- sequence 2: number of correct digits roughly doubles at each step
- sequence 3: number of correct digits gained per step increases slowly (from $0.5$ initially to $2$ near the end)

# Bisection method

$f$ : **R** $\rightarrow$ **R**, continuous




if $f(l)f(u) < 0$, then the interval $[l, u]$ contains at least one zero




---

**given** $l$, $u$ with $l < u$ and $f(l)f(u) < 0$; a required tolerance $\epsilon > 0$
**repeat**

1. $x := (l + u)/2$.
2. Compute $f(x)$.
3. **if** $f(x) = 0$, **return** $x$.
4. **if** $f(x)f(l) < 0$, $u := x$, **else**, $l := x$.

**until** $u - l \leq \epsilon$

---

one function evaluation per iteration

## convergence rate

- $u^{(k)} - l^{(k)}$ measures our uncertainty in localizing a zero $x^\star$:

$$|x^{(k)} - x^\star| \leq u^{(k)} - l^{(k)}$$

- uncertainty is halved at each iteration:

$$\begin{aligned} u^{(k)} - l^{(k)} &= \left(\frac{1}{2}\right)^k (u^{(0)} - l^{(0)}) \\ |x^{(k)} - x^\star| &\leq \left(\frac{1}{2}\right)^k (u^{(0)} - l^{(0)}) \end{aligned}$$

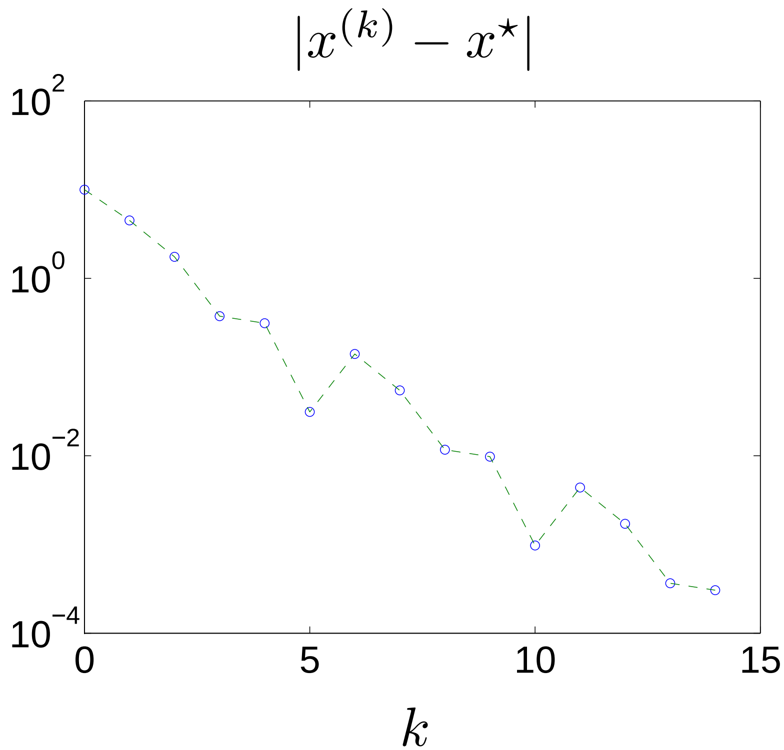
  $i.e.$, R-linear convergence with $c = 1/2$, $M = u^{(0)} - l^{(0)}$

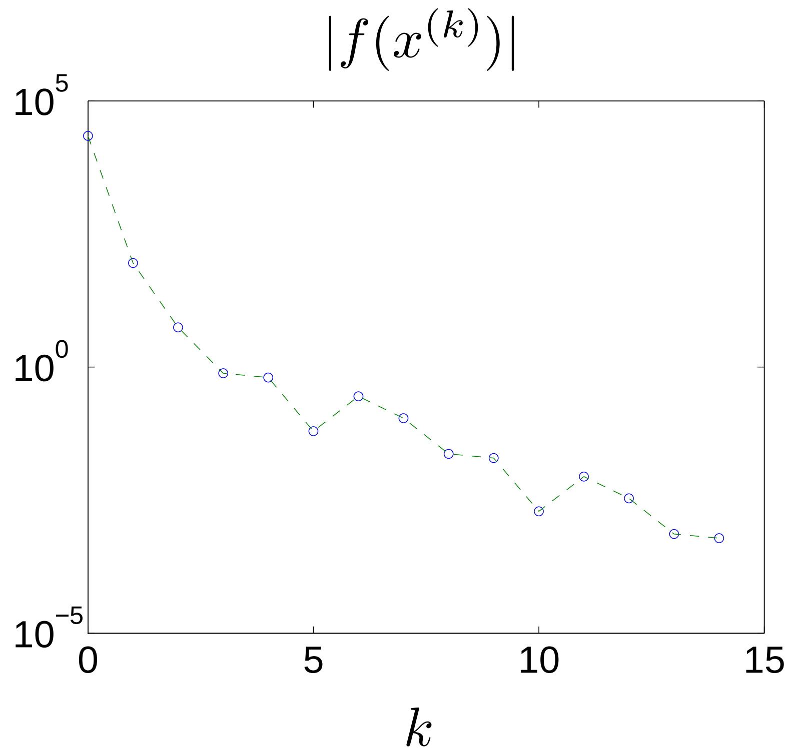
- number of iterations required for $u^{(k)} - l^{(k)} \leq \epsilon$:

$$\log_2 \frac{u^{(0)} - l^{(0)}}{\epsilon}$$

**example:** $f(x) = e^x - e^{-x}$

- unique zero $x^\star = 0$
- start bisection method with $l = -1$, $u = 21$



# Newton's method

$f$ : $\mathbf{R} \to \mathbf{R}$, differentiable

---

**given** initial $x$, required tolerance $\epsilon > 0$
**repeat**

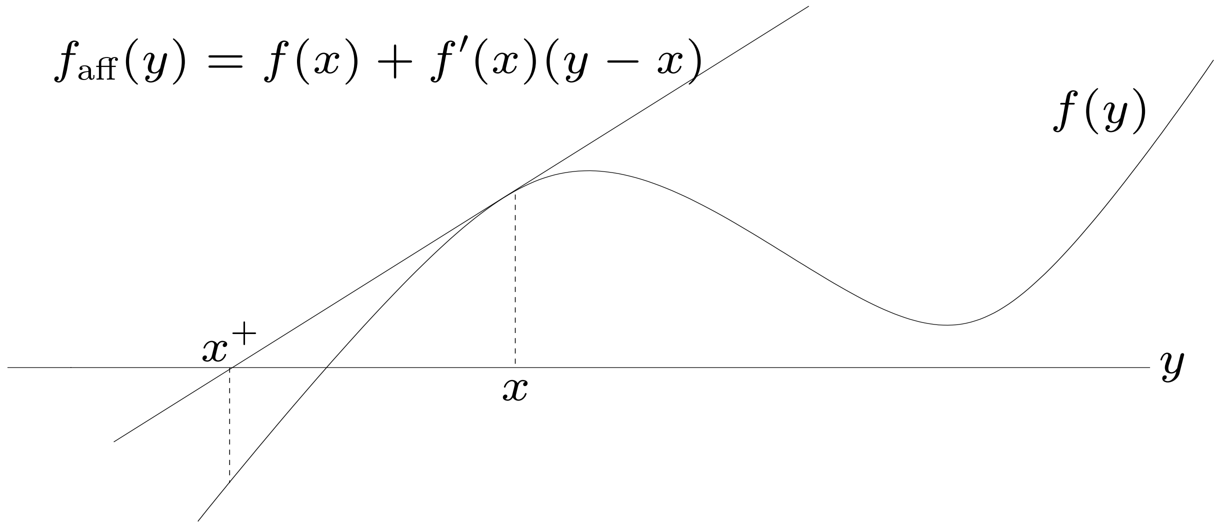
1. Compute $f(x)$ and $f'(x)$.
2. **if** $|f(x)| \leq \epsilon$, **return** $x$.
3. $x := x - f(x)/f'(x)$.

**until** maximum number of iterations is exceeded.

---

- each iteration requires one evaluation of $f$ and $f'$
- there exist other (more sophisticated) stopping criteria
- we assume $f'(x^{(k)}) \neq 0$, all $k$

**interpretation** (with notation $x = x^{(k)}$, $x^+ = x^{(k+1)}$)




- make affine approximation of $f$ around $x$ using Taylor series expansion:

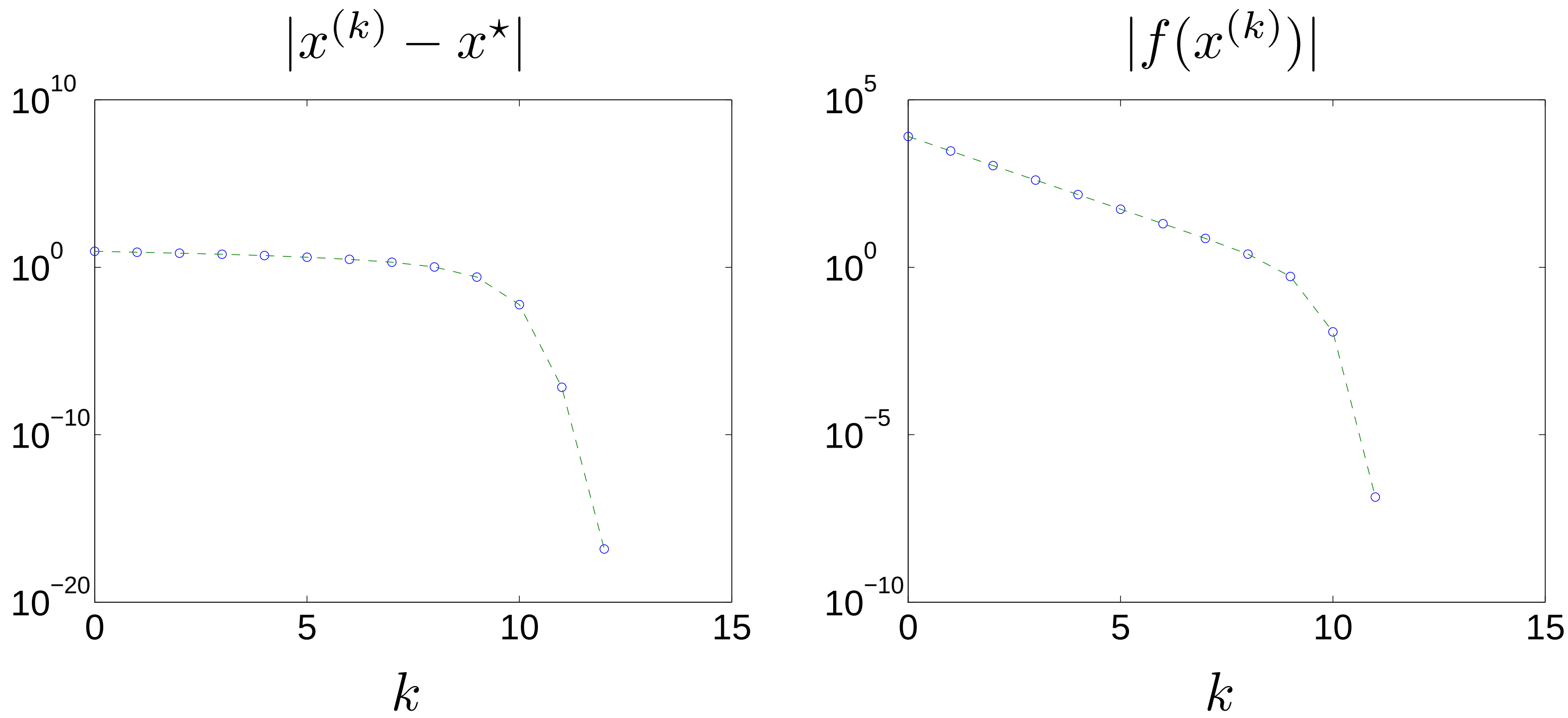
$$f_{\text{aff}}(y) = f(x) + f'(x)(y - x)$$

- solve the linearized equation $f_{\text{aff}}(y) = 0$ and take the solution $y$ as $x^+$:

$$x^+ = x - f(x)/f'(x)$$

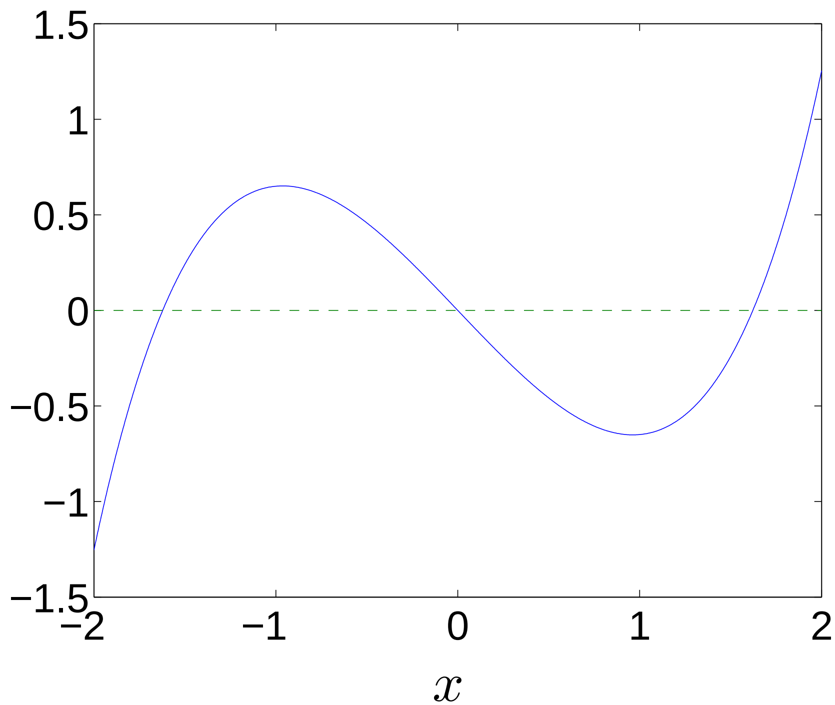
# Examples

- $f(x) = e^x - e^{-x}$, start at $x^{(0)} = 10$




asymptotic convergence is much faster than bisection method

- $f(x) = e^x - e^{-x} - 3x$

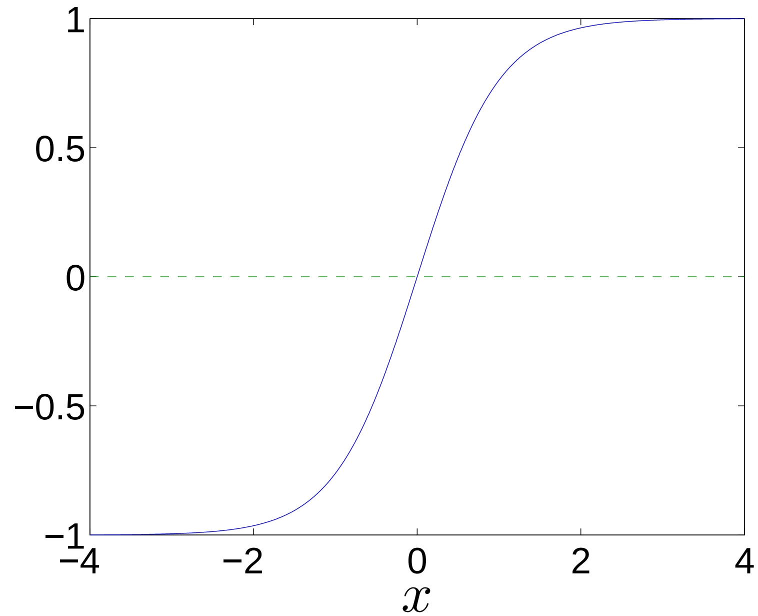



  - start at $x^{(0)} = -1$: converges to $x = -1.62$
  - start at $x^{(0)} = -0.8$: converges to $x = 1.62$
  - start at $x^{(0)} = -0.7$: converges to $x = 0$

converges to a different solution depending on the starting point

- $f(x) = \dfrac{e^x - e^{-x}}{e^x + e^{-x}}$ (unique root at $x = 0$)
  - start at $x^{(0)} = 0.9$:

$$\begin{aligned} x^{(1)} &= -5.7\ 10^{-1} \\ x^{(2)} &= 1.3\ 10^{-1} \\ x^{(3)} &= -1.6\ 10^{-3} \\ x^{(4)} &= 2.5\ 10^{-9} \\ x^{(5)} &= -3.0\ 10^{-17} \end{aligned}$$




converges very rapidly

  - start at $x^{(0)} = 1.1$:

$$x^{(1)} = -1.1\ 10^0, \qquad x^{(2)} = 1.2\ 10^0, \qquad x^{(3)} = -1.7\ 10^0,$$
$$x^{(4)} = 5.7\ 10^0, \qquad x^{(5)} = -2.3\ 10^4$$

does not converge

## conclusion

- Newton's method works very well if we start near a solution
- it may not work at all if we start too far from a solution
- if there are multiple solutions, it may converge to a different solution depending on the starting point; it does not necessarily converge to the solution closest to the starting point

# Secant method

$f$ : **R** $\to$ **R**, continuous

---

**given** two initial points $x$, $x_{\text{prev}}$, required tolerance $\epsilon > 0$
**repeat**

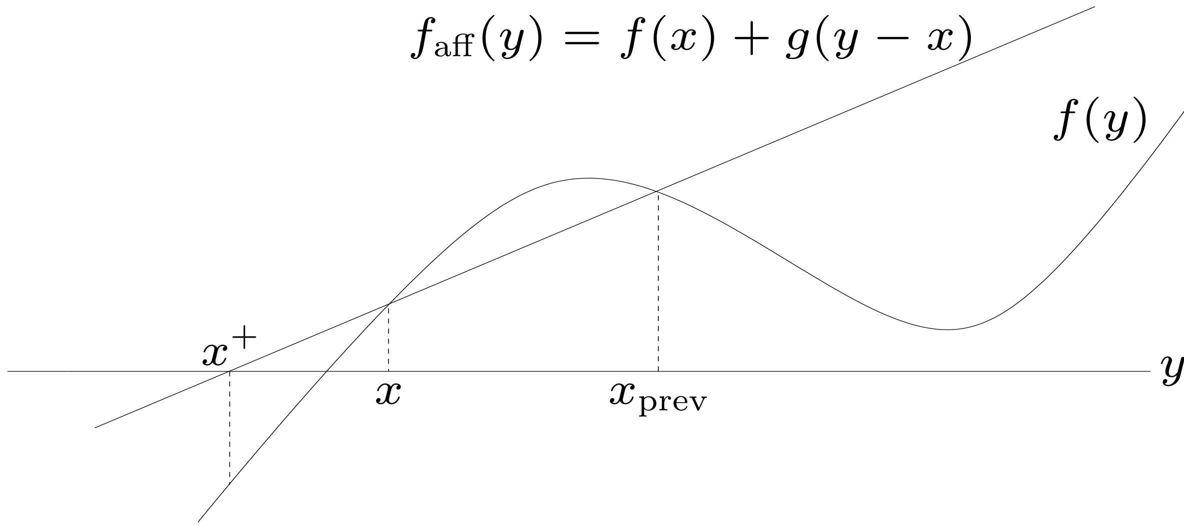
1. Compute $f(x)$
2. **if** $|f(x)| \leq \epsilon$, **return** $x$.
3. $g := (f(x) - f(x_{\text{prev}}))/(x - x_{\text{prev}})$.
4. $x_{\text{prev}} := x$.
5. $x := x - f(x)/g$.

**until** maximum number of iterations is exceeded.

---

- first iteration requires two evaluations of $f$ (at $x$ and $x_{\text{prev}}$)
- subsequent iterations require one evaluation (at $x$)
- we assume $g \neq 0$

**interpretation** (with notation: $x = x^{(k)}$, $x^+ = x^{(k+1)}$, $x_{\text{prev}} = x_{\text{prev}}^{(k)}$)




- affine approximation $f_{\text{aff}}$ with $f_{\text{aff}}(x) = f(x)$, $f_{\text{aff}}(x_{\text{prev}}) = f(x_{\text{prev}})$:

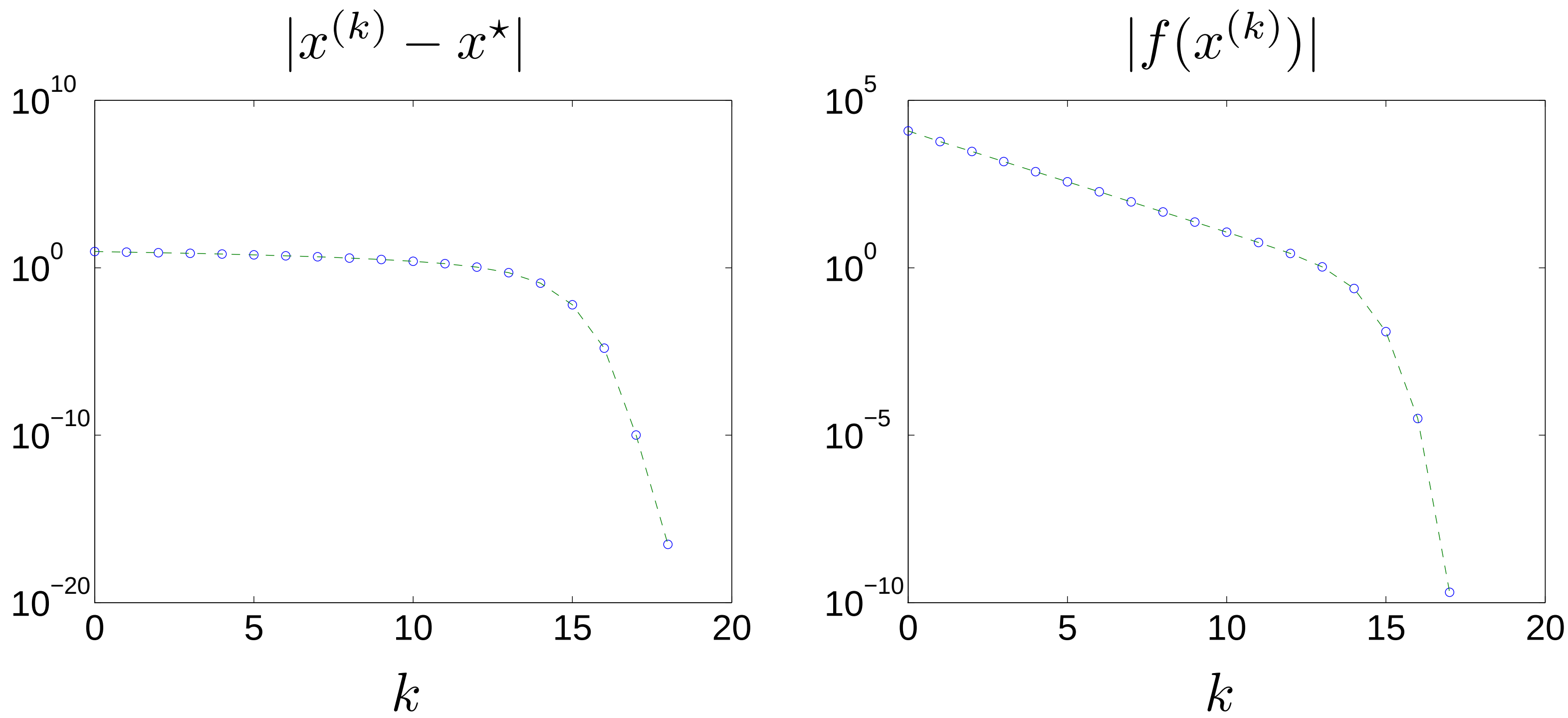
$$f_{\text{aff}}(y) = f(x) + g(y - x) \quad \text{with } g = \frac{f(x) - f(x_{\text{prev}})}{x - x_{\text{prev}}}$$

- solve linear equation $f_{\text{aff}}(y) = 0$ and take the solution as new iterate $x^+$:

$$x^+ = x - f(x)/g$$

# Examples

- $f(x) = e^x - e^{-x}$, start at $x^{(0)} = 10$, $x^{(0)}_{\text{prev}} = 11$




  fast asymptotic convergence, but slower than Newton method

- other examples: secant method works well if we start near a solution; may not converge otherwise

# Convergence of Newton and secant methods

**Newton method**: if $f'(x^\star) \neq 0$ and $x^{(0)}$ is sufficiently close to $x^\star$, then Newton's method converges and there exists a $c > 0$ such that

$$|x^{(k+1)} - x^\star| \leq c\,|x^{(k)} - x^\star|^2$$

$i.e.$, quadratic convergence

**secant method:** if $f'(x^\star) \neq 0$ and $x^{(0)}$ is sufficiently close to $x^\star$, then the secant method converges and there exists a $c > 0$ such that

$$|x^{(k+1)} - x^\star| \leq c\,|x^{(k)} - x^\star|^r$$

where $r = (1+\sqrt{5})/2 \approx 1.6$

$i.e.$, superlinear convergence

# Summary

**bisection method**

- does not require derivatives
- user must provide initial interval $[l, u]$ with $f(l)f(u) < 0$
- R-linear convergence

**Newton's method**

- requires derivatives
- user must provide starting point near a solution
- quadratic convergence

**secant method**

- does not require derivatives
- user must provide two starting points near a solution
- superlinear convergence