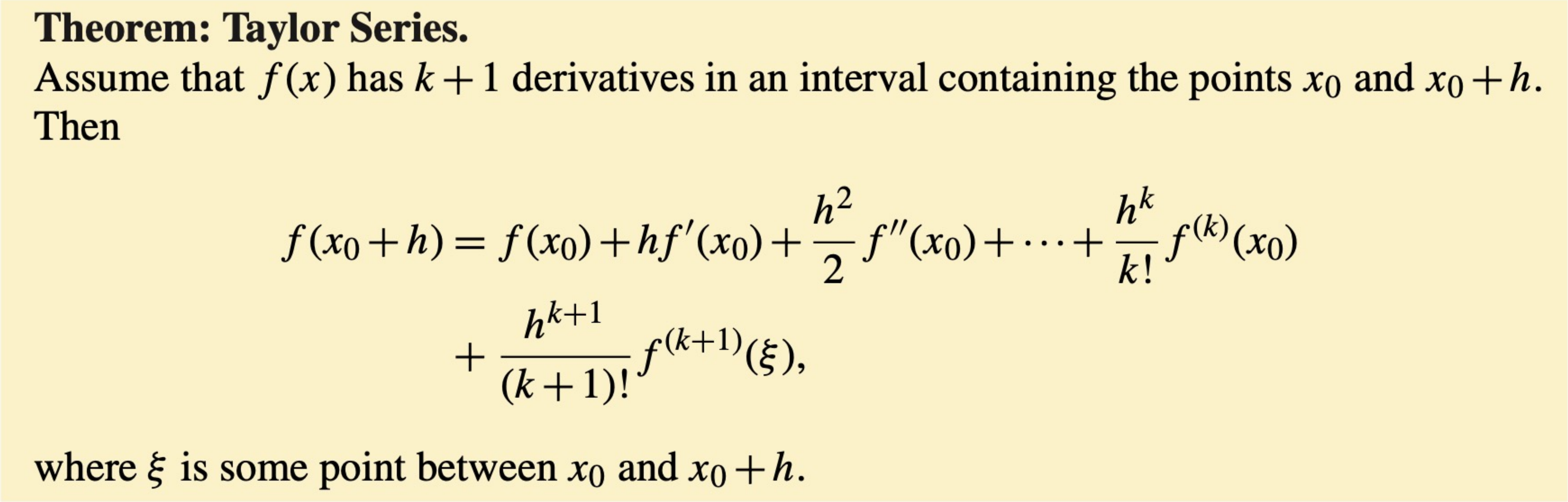
BLG 202 E Numerical Methods

Recitation 1

Tuesday, April 27, 2021

Meral Kuyucu

CHP 1



Theorem: Taylor Series.
Assume that $f(x)$ has $k+1$ derivatives in an interval containing the points x_0 and x_0+h. Then

$$f(x_0+h) = f(x_0) + hf'(x_0) + \frac{h^2}{2} f''(x_0) + \cdots + \frac{h^k}{k!} f^{(k)}(x_0) + \frac{h^{k+1}}{(k+1)!} f^{(k+1)}(\xi),$$

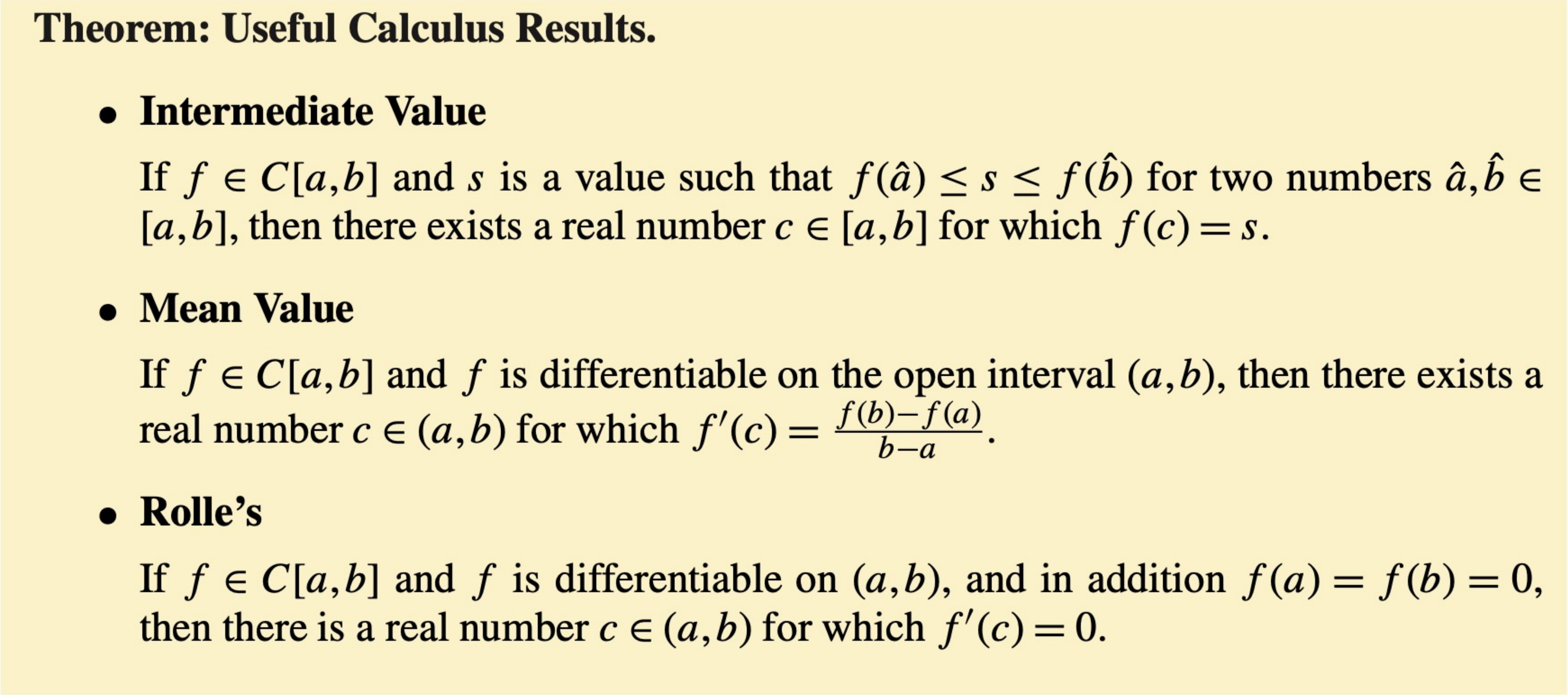
where ξ is some point between x_0 and x_0+h.

Taylor Series Example

- Find the Taylor approximation of f(x) = sin(x) about the point $x = \frac{\pi}{4}$

$$
\begin{aligned}
f(x) &= \sin(x) & &\Longrightarrow & f\left(\frac{\pi}{4}\right) &= \frac{\sqrt{2}}{2} \\
f'(x) &= \cos(x) & &\Longrightarrow & f'\left(\frac{\pi}{4}\right) &= \frac{\sqrt{2}}{2} \\
f''(x) &= -\sin(x) & &\Longrightarrow & f''\left(\frac{\pi}{4}\right) &= -\frac{\sqrt{2}}{2} \\
f'''(x) &= -\cos(x) & &\Longrightarrow & f'''\left(\frac{\pi}{4}\right) &= -\frac{\sqrt{2}}{2} \\
f^{(4)}(x) &= \sin(x) & &\Longrightarrow & f^{(4)}\left(\frac{\pi}{4}\right) &= \frac{\sqrt{2}}{2}
\end{aligned}
$$

$$
p_4(x) = \frac{\sqrt{2}}{2}\left\{1 + \left(x - \frac{\pi}{4}\right) - \frac{1}{2}\left(x - \frac{\pi}{4}\right)^2 - \frac{1}{6}\left(x - \frac{\pi}{4}\right)^3 + \frac{1}{24}\left(x - \frac{\pi}{4}\right)^4\right\}
$$



Theorem: Useful Calculus Results.

- **Intermediate Value**

  If $f \in C[a,b]$ and s is a value such that $f(\hat{a}) \leq s \leq f(\hat{b})$ for two numbers $\hat{a}, \hat{b} \in [a,b]$, then there exists a real number $c \in [a,b]$ for which $f(c) = s$.

- **Mean Value**

  If $f \in C[a,b]$ and f is differentiable on the open interval (a,b), then there exists a real number $c \in (a,b)$ for which $f'(c) = \frac{f(b)-f(a)}{b-a}$.

- **Rolle's**

  If $f \in C[a,b]$ and f is differentiable on (a,b), and in addition $f(a) = f(b) = 0$, then there is a real number $c \in (a,b)$ for which $f'(c) = 0$.

CHP 1 Review Questions:

Give a simple example where relative error is a more suitable measure than absolute error, and another example where the absolute error measure is more suitable.

- The absolute error gives how large the error is, while the relative error gives how large the error is relative to the correct value.
- The relative error is often used to compare approximations of numbers of widely differing size.
- For example, approximating the number 1,000 with an absolute error of 3 is, in most applications, much worse than approximating the number 1,000,000 with an absolute error of 3; in the first case the relative error is 0.003 and in the second it is only 0.000003.

CHP 1 Review Questions:

State a major difference between the nature of roundoff errors and discretization errors.

- Discretization errors arise from discretization of continuous processes, such as interpolation, differentiation, and integration.
- Roundoff errors arise because of the finite precision representation of real numbers on any computer, which affects both data representation and computer arithmetic.

CHP 1 Review Questions:

Explain briefly why accumulation of roundoff errors is inevitable when arithmetic operations are performed in a floating point system. Under which circumstances is it tolerable in numerical computations?

- In general it is impossible to prevent linear accumulation, meaning the roundoff error may be proportional to n after n elementary operations such as addition or multiplication of two real numbers.
- However, such an error accumulation is usually acceptable if the linear rate is moderate (i.e., the constant c0 below is not very large). In contrast, exponential growth cannot be tolerated.

$$E_n \simeq c_0 n E_0 \text{ for some constant } c_0 \text{ represents linear growth, and}$$
$$E_n \simeq c_1^n E_0 \text{ for some constant } c_1 > 1 \text{ represents exponential growth.}$$

CHP 1 Review Questions:

Explain the differences between accuracy, efficiency, and robustness as criteria for evaluating an algorithm.

1. Accuracy – Magnitude of Error
2. Efficiency – Use of Resources (Time and Space)
3. Robustness – Algorithm Works Well in All Conditions

CHP 1 Review Questions:

Show that nested evaluation of a polynomial of degree *n* requires only 2*n* elementary operations and hence has O(*n*) complexity.

$$p_n(x) = c_0 + c_1 x + \cdots + c_n x^n.$$

n sums & n multiplications (between x and coefficients)

Powers of x : 0 + 1 + … + n = $(n^2 + n)/2$

Combined: 2n + $(n^2+n)/2$ is $O(n^2)$

$$p_n(x) = (\cdots((c_n x + c_{n-1})x + c_{n-2})x \cdots)x + c_0$$

n nested parentheses

1 addition and 1 multiplication per set of parentheses

Combined: 2n is O(n)

CHP 2

Real Numbers in Decimal

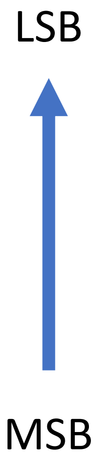
$$\begin{tikzcd}
\frac{69162}{125} \arrow[d] & \text{Fraction} \\
553\frac{37}{125} \arrow[d] & \text{Mixed Number} \\
553.296 \arrow[d] & \text{Decimal Number} \\
553.296 = 5 * 10^2 + 5 * 10^1 + 3 * 10^0 + 2 * 10^{-1} + 9 * 10^{-2} + 6 * 10^{-2} &
\end{tikzcd}$$

Decimal Expansion

Decimal to Binary (Whole Numbers)

- Convert $(11.1875)_{10}$ to base 2 (binary representation)
- $(11)_{10} = (1011)_2$
  - Check:
  - $1 * 2^3 + 0 * 2^2 + 1 * 2^1 + 1 * 2^0 = 8 + 0 + 2 + 1 = 11$

	Q	R
11/2	5	1
5/2	2	1
2/2	1	0
1/2	0	1



LSB

MSB

Decimal to Binary (Fractions)

- Convert $(11.1875)_{10}$ to base 2 (binary representation)
- $(0.1875)_{10} = (0.0011)_2$
  - Check:
  - $0 * 2^{-1} + 0 * 2^{-2} + 1 * 2^{-3} + 1 * 2^{-4} = 0 + 0 + 0.125 + 0.0625 = 0.1875$

	Product (x2)	After Decimal	Before Decimal
0.1875	0.375	0.375	0
0.375	0.75	0.75	0
0.75	1.5	0.5	1
0.5*2	1	0	1

$$\boxed{(11.1875)_{10} = (1101.0011)_2}$$

Chopping Error

$y_{real} = 1.956853$

- Evaluate the polynomial $y = x^3 - 5x^2 + 6x + 0.55$ at $x = 1.37$. Use 3-digit arithmetic with chopping. Evaluate the percent relative error.
  - $y_{chop} = 1.37^3 - 5(1.37^2) + 6(1.37) + 0.55$
  - $y_{chop} = 2.56 - 5(1.87) + 8.22 + 0.55$
  - $y_{chop} = 2.56 - 9.35 + 8.22 + 0.55 = 1.98$
  - $y_{chop} = 1.98$

$$\epsilon_{chop} = \frac{|1.956853 - 1.98|}{|1.956853|} * 100\% = 1.183\%$$

- Repeat but express y as $y = \big((x-5)x + 6\big)x + 0.55$
  - $y_{chop} = ((1.37 - 5)1.37 + 6)1.37 + 0.55$
  - $y_{chop} = ((-3.63)1.37 + 6)1.37 + 0.55$
  - $y_{chop} = (-4.97 + 6)1.37 + 0.55$
  - $y_{chop} = 1.03 * 1.37 + 0.55$
  - $y_{chop} = 1.41 + 0.55$
  - $y_{chop} = 1.96$

$$\epsilon_{chop} = \frac{|1.956853 - 1.96|}{|1.956853|} * 100\% = 0.161\%$$

Rounding Error

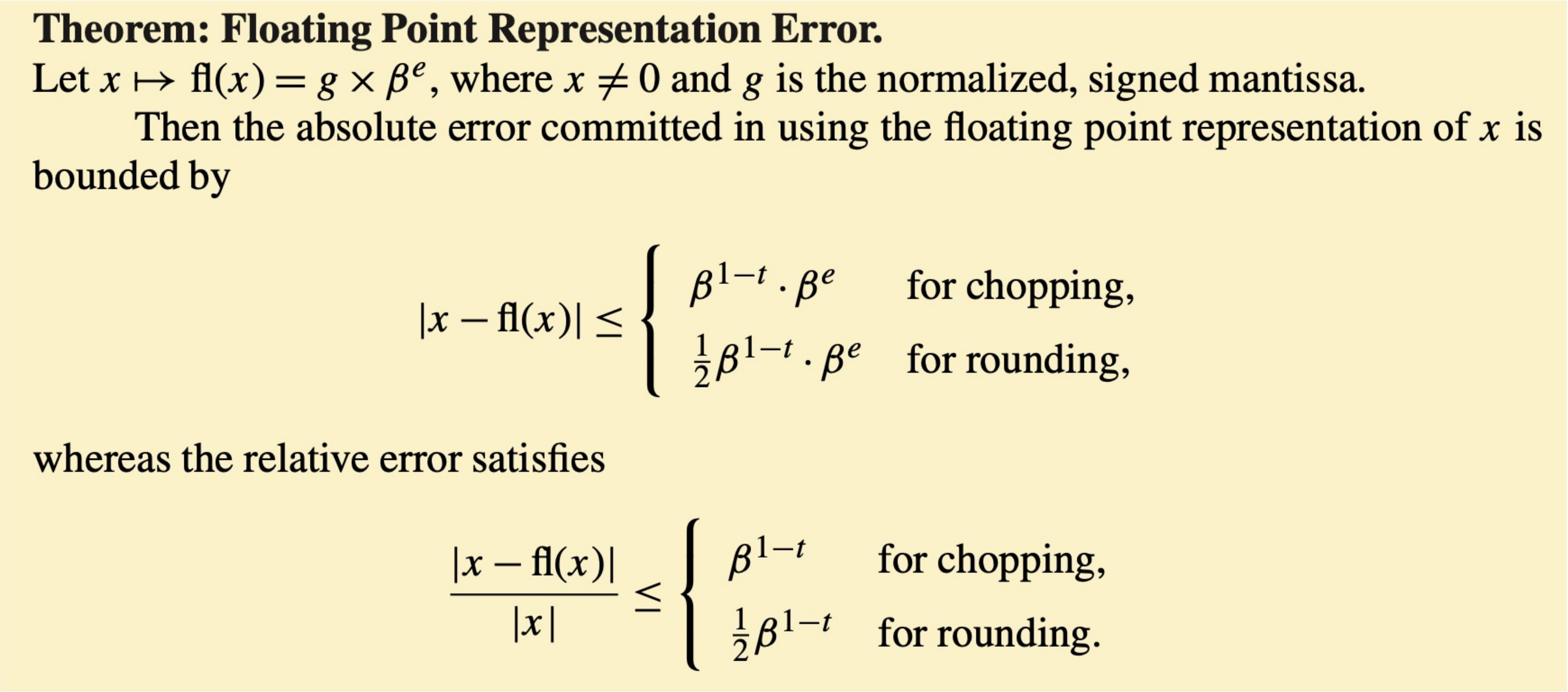
$y_{real} = 1.956853$

- Evaluate the polynomial $y = x^3 - 5x^2 + 6x + 0.55$ at $x = 1.37$. Use 3-digit arithmetic with rounding. Evaluate the percent relative error.
  - $y_{round} = 1.37^3 - 5(1.37^2) + 6(1.37) + 0.55$
  - $y_{round} = 2.58 - 5(1.88) + 8.22 + 0.55$
  - $y_{round} = 2.58 - 9.40 + 8.22 + 0.55 = 1.95$
  - $y_{round} = 1.95$

$$\in_{round} = \frac{|1.956853 - 1.95|}{|1.956853|} * 100\% = 0.35\%$$

- Repeat but express y as $y = \big((x-5)x + 6\big)x + 0.55$
  - $y_{round} = ((1.37 - 5)1.37 + 6)1.37 + 0.55$
  - $y_{round} = ((-3.63)1.37 + 6)1.37 + 0.55$
  - $y_{round} = (-4.97 + 6)1.37 + 0.55$
  - $y_{round} = 1.03 * 1.37 + 0.55$
  - $y_{round} = 1.41 + 0.55$
  - $y_{round} = 1.96$

$$\in_{round} = \frac{|1.956853 - 1.96|}{|1.956853|} * 100\% = 0.161\%$$



Theorem: Floating Point Representation Error.

Let $x \mapsto \mathrm{fl}(x) = g \times \beta^e$, where $x \neq 0$ and g is the normalized, signed mantissa.

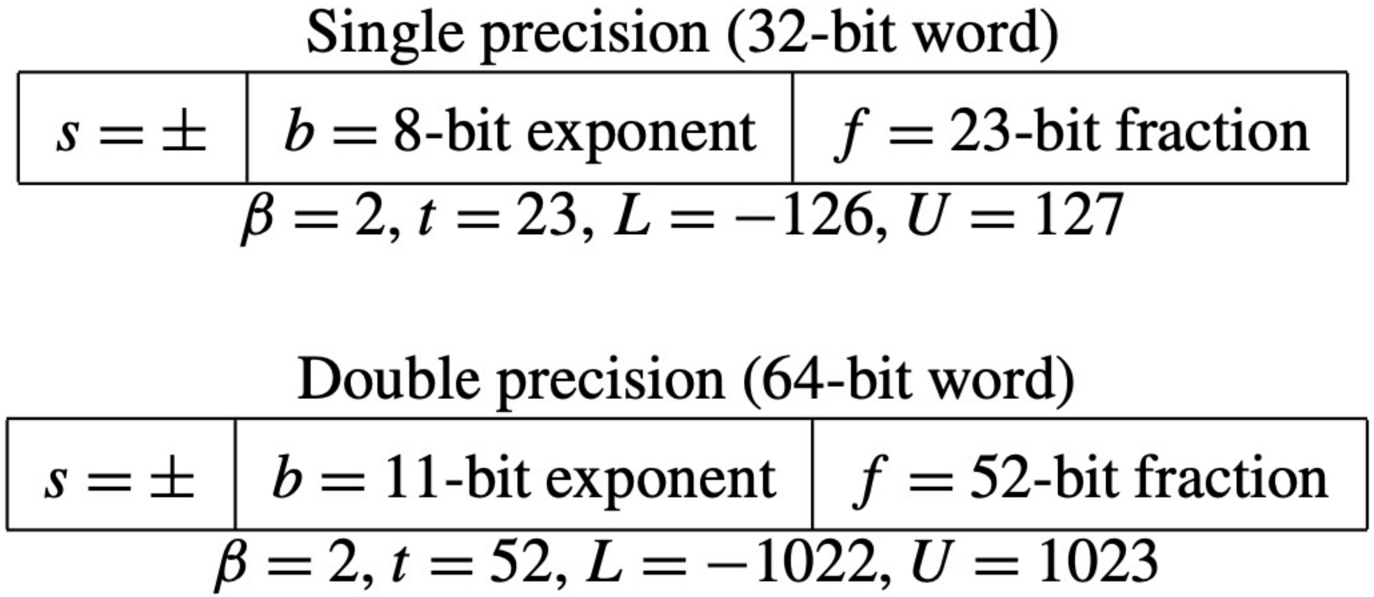
Then the absolute error committed in using the floating point representation of x is bounded by

$$|x - \mathrm{fl}(x)| \leq \begin{cases} \beta^{1-t} \cdot \beta^e & \text{for chopping,} \\ \frac{1}{2}\beta^{1-t} \cdot \beta^e & \text{for rounding,} \end{cases}$$

whereas the relative error satisfies

$$\frac{|x - \mathrm{fl}(x)|}{|x|} \leq \begin{cases} \beta^{1-t} & \text{for chopping,} \\ \frac{1}{2}\beta^{1-t} & \text{for rounding.} \end{cases}$$

IEEE Standard



Single precision (32-bit word)

$s = \pm$	$b =$ 8-bit exponent	$f =$ 23-bit fraction

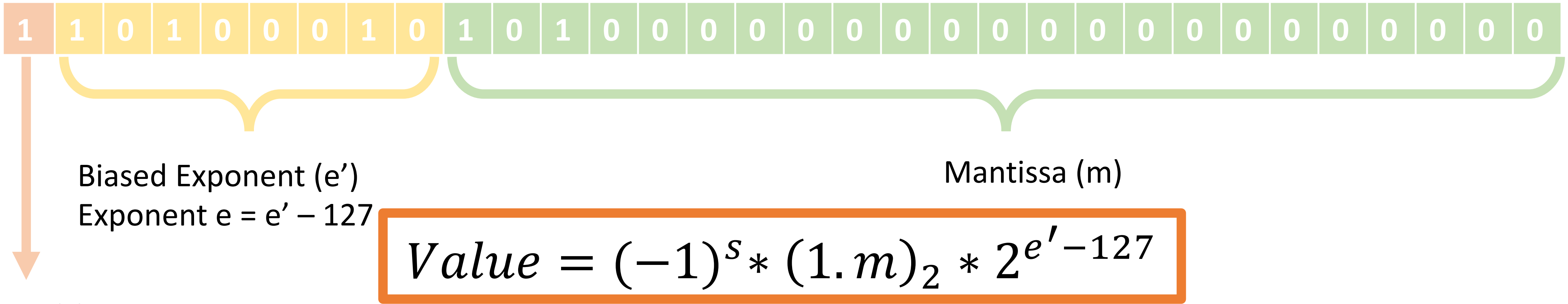
$\beta = 2, t = 23, L = -126, U = 127$

Double precision (64-bit word)

$s = \pm$	$b =$ 11-bit exponent	$f =$ 52-bit fraction

$\beta = 2, t = 52, L = -1022, U = 1023$

IEEE Standard



1	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Biased Exponent (e')
Exponent e = e' – 127

Mantissa (m)

Sign Bit (s)
0 = + & 1 = -

$$\boxed{Value = (-1)^{s} * (1.m)_2 * 2^{e'-127}}$$

$$Value = (-1)^{1} * (1.101)_2 * 2^{(10100010)_2 - 127}$$
$$Value = (-1) * (1.625)_{10} * 2^{(162)_{10} - 127}$$
$$Value = (-1) * (1.625) * 2^{35}$$
$$Value = -55834574848$$

In Scientific Notation Chopped at t=5 digits:

$$Value = -5.5834x10^{10}$$

CHP 2 Review Questions:

What is a normalized floating point number and what is the purpose of normalization?

- Normalization is the process of moving the binary point so that the first digit after the point is a significant digit. This maximizes precision in a given number of bits. To maximize the precision of a positive number you should have a mantissa with no leading zeros.

CHP 2 Review Questions:

Define overflow and underflow. Why is the former considered more damaging than the latter?

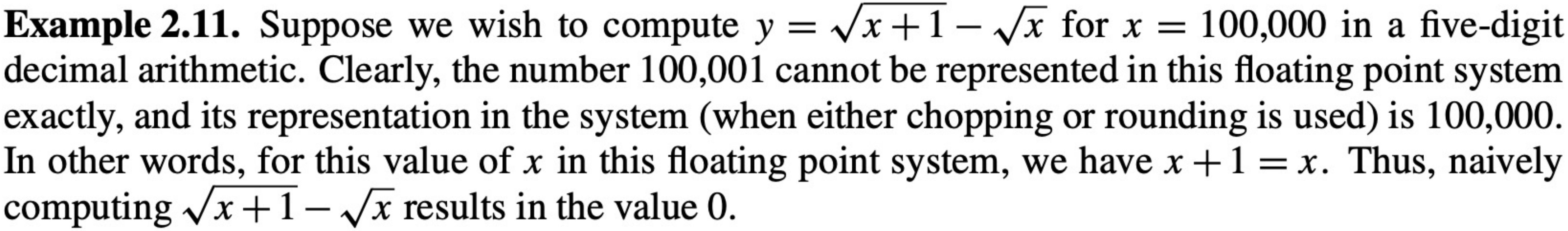
- An overflow is obtained when a number is too large to fit into the floating point system in use, i.e., when $e > U$.
- An underflow is obtained when $e < L$ (number is too close to 0).
- When overflow occurs in the course of a calculation, this is generally fatal. But underflow is nonfatal: the system usually sets the number to 0 and continues.

CHP 2 Review Questions:

What is a cancellation error? Give an example of an application where it arises in a natural way.

- Cancellation error occurs when two nearly equal numbers are subtracted from one another.

$$f'(x_0) \approx \frac{f(x_0+h) - f(x_0)}{h}$$



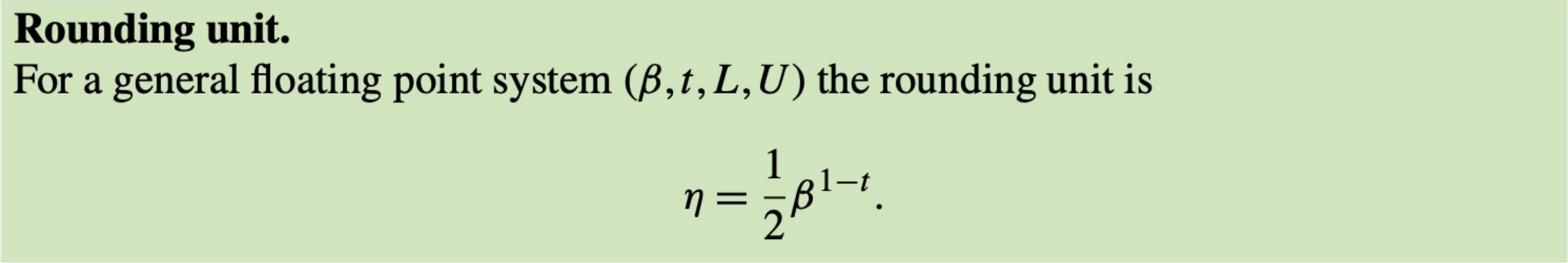
Example 2.11. Suppose we wish to compute $y = \sqrt{x+1} - \sqrt{x}$ for $x = 100{,}000$ in a five-digit decimal arithmetic. Clearly, the number 100,001 cannot be represented in this floating point system exactly, and its representation in the system (when either chopping or rounding is used) is 100,000. In other words, for this value of x in this floating point system, we have $x+1 = x$. Thus, naively computing $\sqrt{x+1} - \sqrt{x}$ results in the value 0.

CHP 2 Review Questions:

Define rounding unit and explain its importance.

What is the rounding unit for base β = 2 and t = 52 digits?

$$\eta = \frac{1}{2} 2^{1-52} = 2^{-52}$$



Rounding unit.

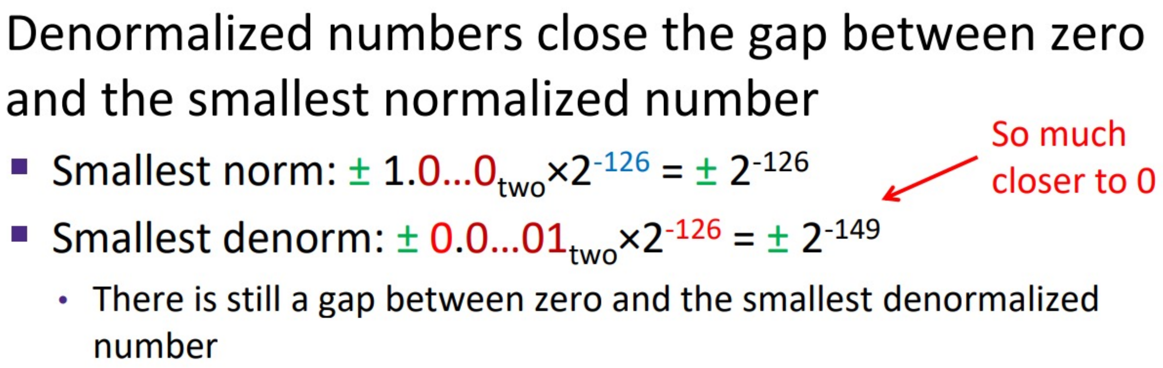
For a general floating point system (β, t, L, U) the rounding unit is

$$\eta = \frac{1}{2}\beta^{1-t}.$$

CHP 2 Review Questions:

Under what circumstances could nonnormalized floating point numbers be desirable?

- The purpose of having subnormal/denormalized numbers is to smooth the gap between the smallest normal number and zero.
- It is very important to realize that subnormal numbers are represented with less precision than normal numbers. In fact, they are trading reduced precision for their smaller size.



Denormalized numbers close the gap between zero and the smallest normalized number

- Smallest norm: $\pm\ 1.0\ldots0_{two}\times2^{-126} = \pm\ 2^{-126}$
- Smallest denorm: $\pm\ 0.0\ldots01_{two}\times2^{-126} = \pm\ 2^{-149}$
  - There is still a gap between zero and the smallest denormalized number

So much closer to 0

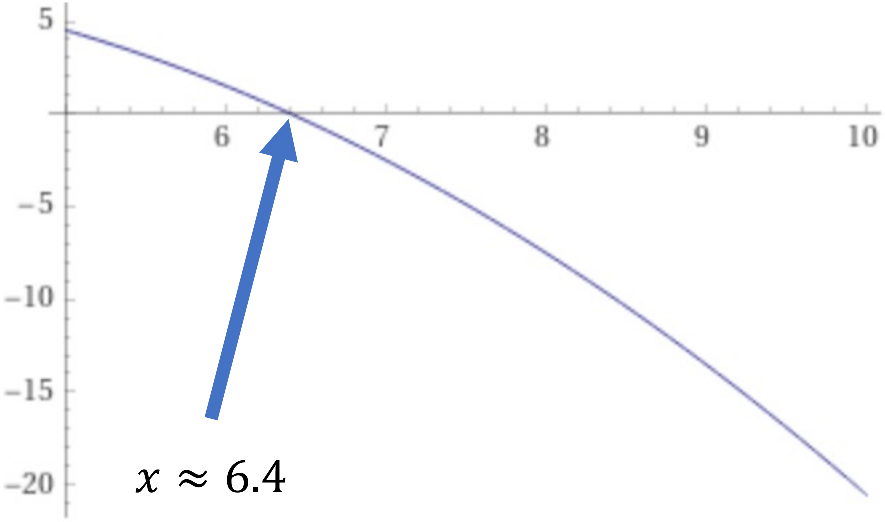
CHP 3

Bisection Method

1. Choose lower x_l and upper x_u guesses for the root such that the function changes sign over the interval.
   - To check: $f(x_l) * f(x_u) < 0$
2. Determine estimate of the root x_r by $x_r = \frac{x_l + x_u}{2}$
3. Make the following evaluations to determine in which subinterval the root lies:
   - If $f(x_l) * f(x_r) < 0$: Root is in **<u>lower</u>** subinterval ($x_{r*} \in [x_l, x_r]$)
     - Set $x_u = x_r$ & Return to step 2.
   - If $f(x_l) * f(x_r) > 0$: Root is in **<u>upper</u>** subinterval ($x_{r*} \in [x_r, x_u]$)
     - Set $x_l = x_r$ & Return to step 2.
   - If $f(x_l) * f(x_r) = 0$: **Root = $\boldsymbol{x_r}$**
     - Terminate

Bisection Method Example

- Determine the real root of $\boldsymbol{f(x) = -0.5x^2 + 2.5x + 4.5}$ using bisection to locate the root.
- Employ initial guesses of $\boldsymbol{x_l = 5}$, $\boldsymbol{x_u = 10}$ and $\boldsymbol{atol = 0.1}$
- Graphical Representation:



Number of Iterations:

$$\frac{b-a}{2} \cdot 2^{-n} \leq \texttt{atol}$$

$$n = \left\lceil \log_2\left(\frac{b-a}{2\,\texttt{atol}}\right) \right\rceil$$

- Can predetermine number of iterations using atol:
- $\frac{(b-a)}{2} * 2^{-n} \leq atol$
- $\frac{(10-5)}{2} * 2^{-n} \leq 0.1$
- $(5) * 2^{-n} \leq 2 * 0.1$
- $2^{-n} \leq \frac{0.2}{5}$
- $\log_2 2^{-n} \leq \log_2 0.04$
- $-n \leq -4.64$
- $n \geq 4.64$
- $n \geq 5$

We need 5 iterations to find the root for atol = 0.1

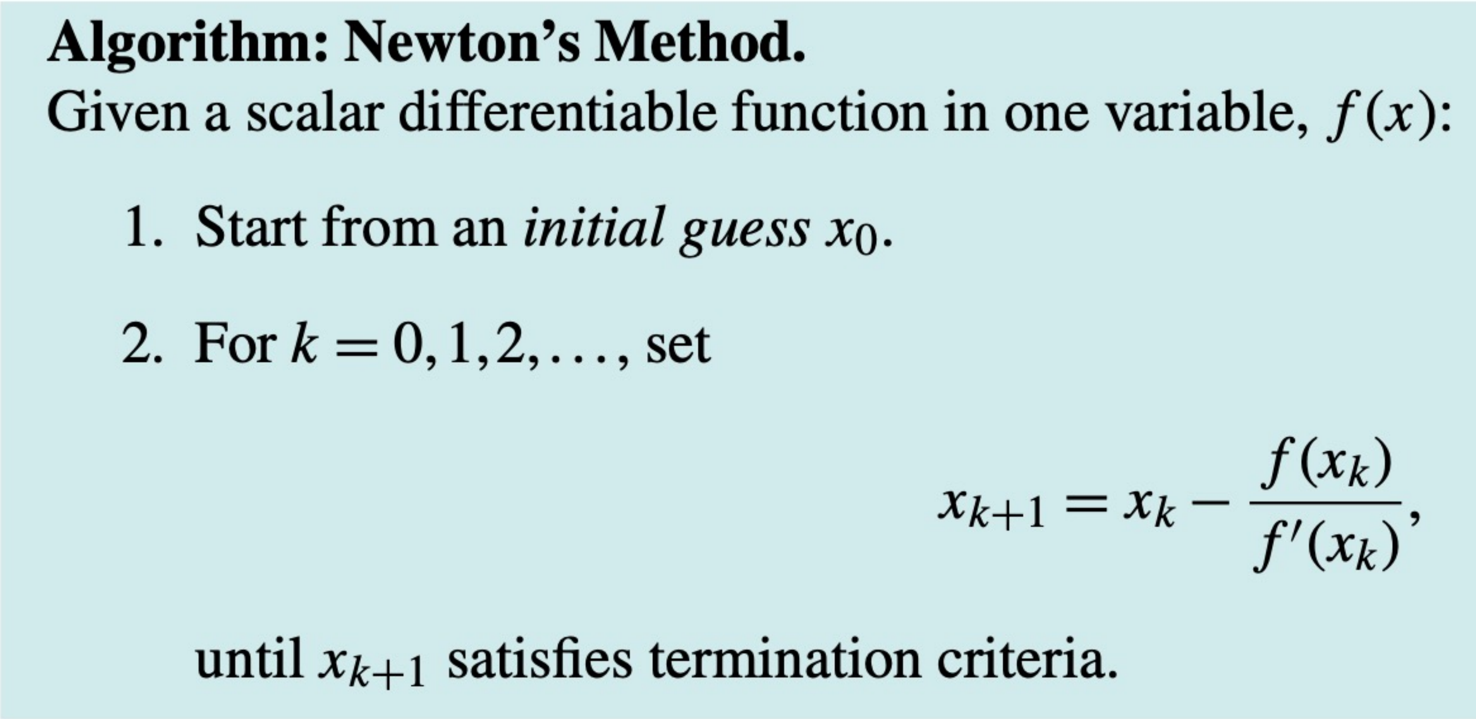
Bisection Method Solution

Iteration	x_l	x_r	x_u	$f(x_l)$	$f(x_r)$	$f(x_u)$	$f(x_l)*f(x_r)$	Sign
1	5	7.5	10	4.5	-4.875	-20.5	-21.9375	Neg
2	5	6.25	7.5	4.5	0.59375	-4.875	2.671875	Pos
3	6.25	6.875	7.5	0.59375	-1.945313	-4.875	-1.15503	Neg
4	6.25	6.5625	6.875	0.59375	-0.626953	-1.94531	-0.37225	Neg
5	6.25	6.40625	6.5625	0.59375	-0.004395	-0.626953	-0.00261	Neg

$$\boxed{|x^* - x_n| \leq \texttt{atol}}$$

$$\boxed{\begin{array}{c} |6.40512 - 6.40625| \leq 0.1 \\ 0.00113 \leq 0.1 \end{array}}$$

Newton’s Method

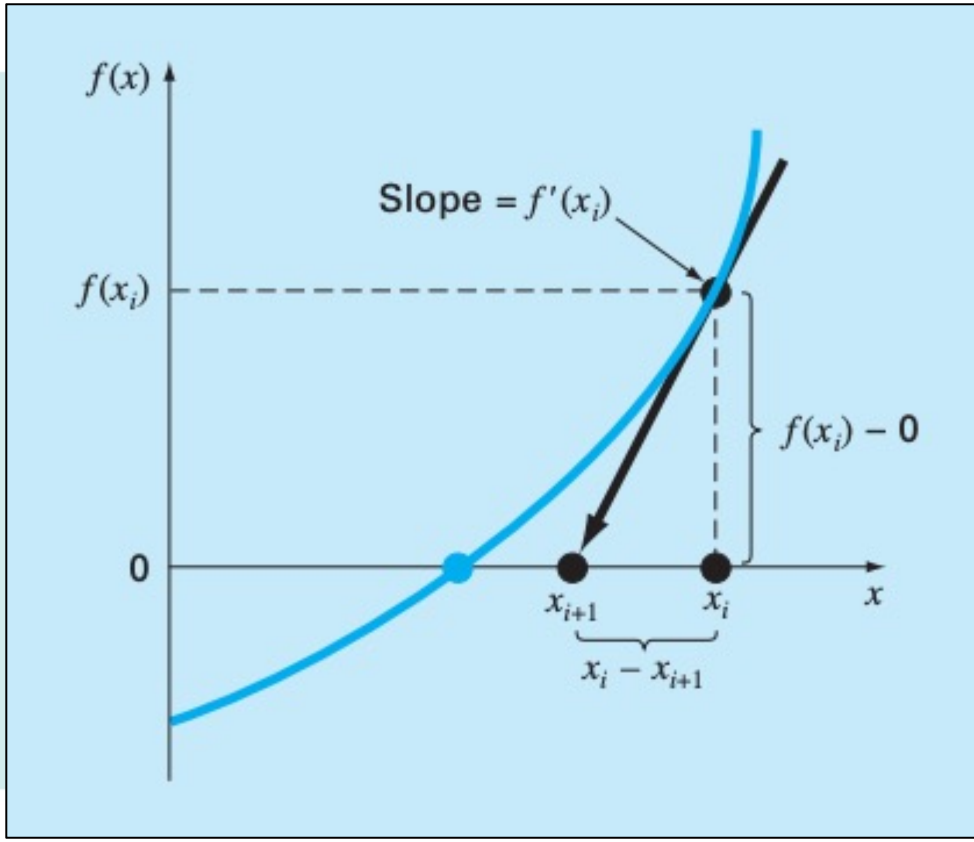


Algorithm: Newton’s Method.
Given a scalar differentiable function in one variable, $f(x)$:

1. Start from an *initial guess* x_0.
2. For $k = 0, 1, 2, \ldots,$ set

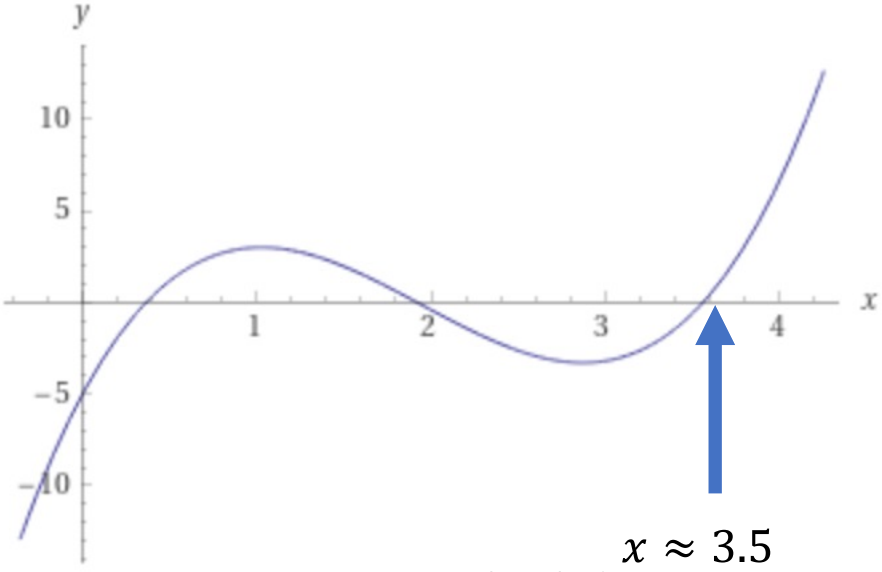
$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)},$$

until x_{k+1} satisfies termination criteria.



Newton’s Method Example

- Find the highest real root of $\boldsymbol{f(x) = 2x^3 - 11.7x^2 + 17.7x - 5}$
- Compute the approximate percent relative errors for your solutions.
- Number of Iterations $\boldsymbol{(n) : 3}$
- Initial Guess $\boldsymbol{(x_0) : 3}$
- Graphical Represen



$x \approx 3.5$

Newton’s Method Solution

- $f(x) = 2x^3 - 11.7x^2 + 17.7x - 5$
- $f'(x) = 6x^2 - 23.4x + 17.7$

- $x_0 = 3$
- I1: $x_1 = 3 - \frac{f(3)}{f'(3)} = 3 - \frac{2(3^3)-11.7(3^2)+17.7(3)-5}{6(3^2)-23.4(3)+17.7} = 3 - \frac{-3.2}{1.5} = \frac{77}{15} \approx 5.1333$
- I2: $x_2 = 5.1333 - \frac{f(5.1333)}{f'(5.1333)} = \frac{2(5.1333^3)-11.7(5.1333^2)+17.7(5.1333)-5}{6(5.1333^2)-23.4(5.1333)+17.7} \approx 4.2697$
- I3: $x_2 = 4.2697 - \frac{f(4.2697)}{f'(4.2697)} = \frac{2(4.2697^3)-11.7(4.2697^2)+17.7(4.2697)-5}{6(4.2697^2)-23.4(4.2697)+17.7} \approx 3.7929$

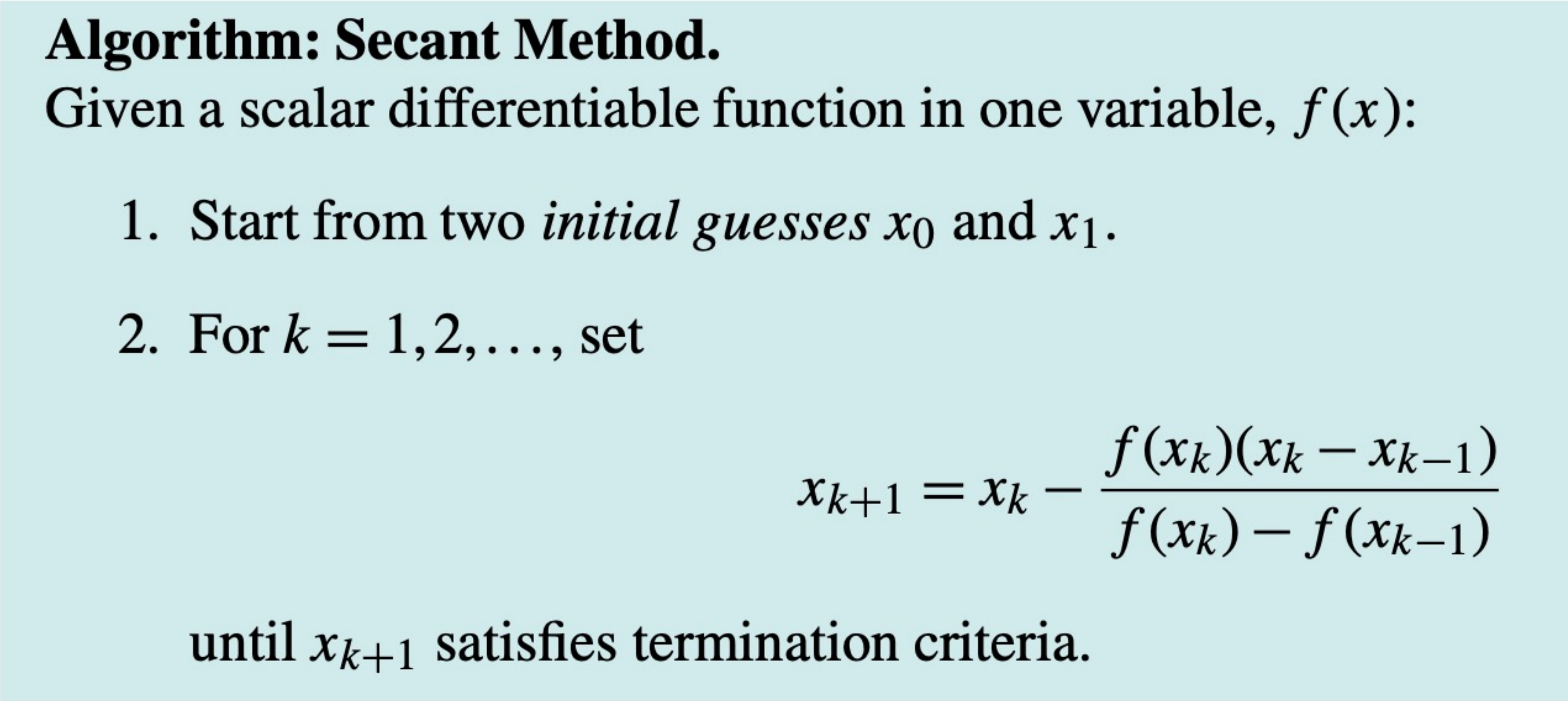
$$\varepsilon = \frac{5.1333 - 3}{5.133} = 0.416$$

$$\varepsilon = \frac{4.2697 - 5.1333}{4.2697} = -0.202$$

$$\varepsilon = \frac{3.7929 - 4.2697}{3.7929} = -0.126$$

$$\boxed{f(x) = 0 \implies x \approx 3.7929}$$

Secant Method

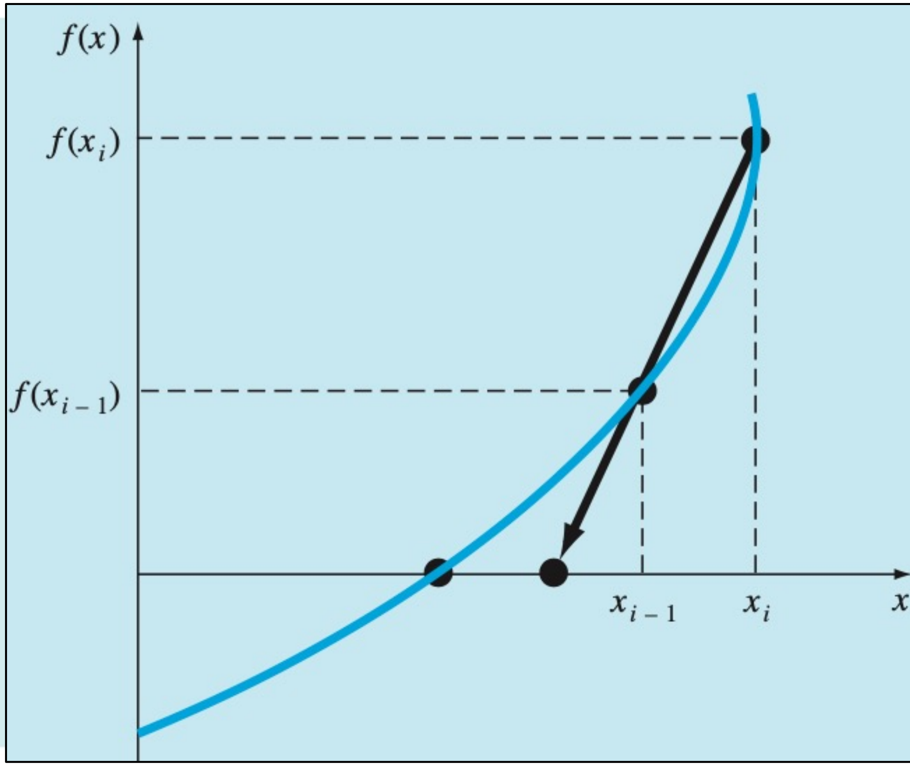


Algorithm: Secant Method.
Given a scalar differentiable function in one variable, $f(x)$:

1. Start from two *initial guesses* x_0 and x_1.
2. For $k = 1, 2, \ldots,$ set

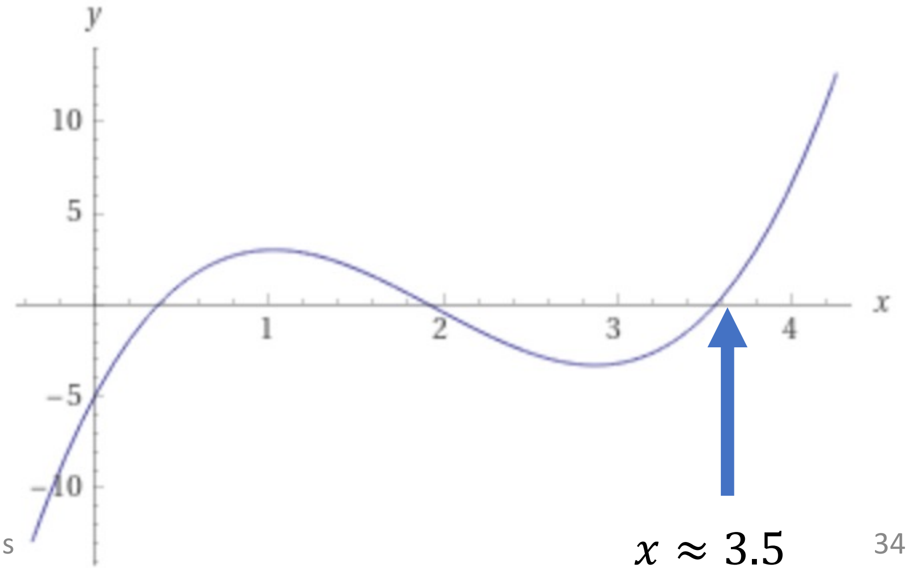
$$x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k) - f(x_{k-1})}$$

until x_{k+1} satisfies termination criteria.



Secant Method Example

- Find the highest real root of $\boldsymbol{f(x) = 2x^3 - 11.7x^2 + 17.7x - 5}$
- Compute the approximate percent relative errors for your solutions.
- Number of Iterations $\boldsymbol{(n)}$ **: 3**
- Initial Guesses $\boldsymbol{(x_{-1})}$ **: 3 &** $\boldsymbol{(x_0)}$ **: 4**
- Graphical Representation:



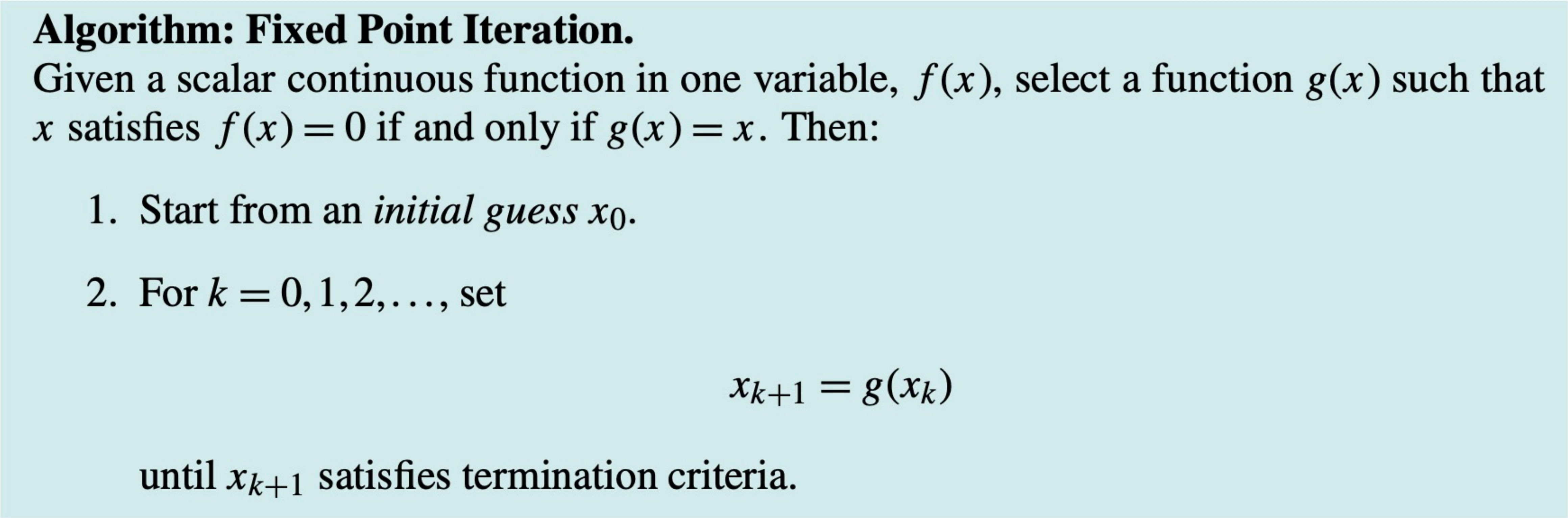
$x \approx 3.5$

Secant Method Solution

- $f(x) = 2x^3 - 11.7x^2 + 17.7x - 5$ & $f(x) \approx \frac{f(x_k)-f(x_{k-1})}{(x_k - x_{k-1})}$
- $x_{k+1} = x_k - \frac{f(x_k)(x_k - x_{k-1})}{f(x_k)-f(x_{k-1})}$
- $x_{-1} = 3$ & $x_0 = 4$
- I1: $x_1 = x_0 - \frac{f(x_0)(x_0 - x_{-1})}{f(x_0)-f(x_{-1})} = 4 - \frac{(6.6)(4-3)}{(6.6)-(-3.2)} = 3.3265$ $\quad \varepsilon = \frac{3.3265 - 4}{3.3265} = -0.202$
- I2: $x_2 = x_1 - \frac{f(x_1)(x_1 - x_0)}{f(x_1)-f(x_0)} = 3.3265 - \frac{(-1.9690)(3.3265-4)}{(-1.9690)-(6.6)} = 3.4813$ $\quad \varepsilon = \frac{3.4813 - 3.3265}{3.4813} = 0.044$
- I3: $x_3 = x_2 - \frac{f(x_2)(x_2 - x_1)}{f(x_2)-f(x_1)} = 3.4813 - \frac{(-0.7957)(3.4813-3.3265)}{(-0.7957)-(-1.9690)} = 3.5863$ $\quad \varepsilon = \frac{3.5863 - 3.4813}{3.5863} = 0.029$

$$\boxed{f(x) = 0 \implies x \approx 3.5863}$$

Fixed Point Iteration



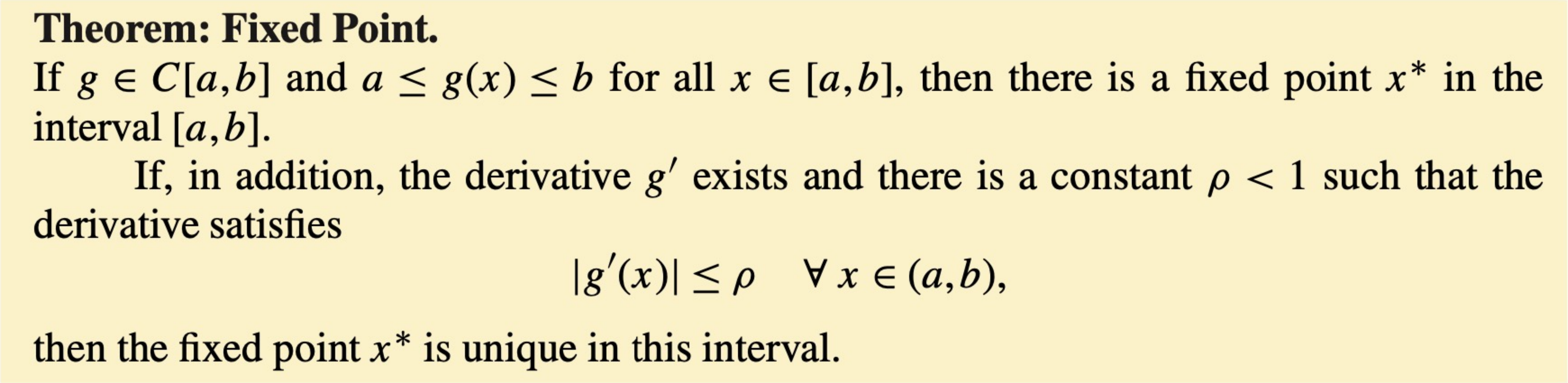
Algorithm: Fixed Point Iteration.
Given a scalar continuous function in one variable, $f(x)$, select a function $g(x)$ such that x satisfies $f(x) = 0$ if and only if $g(x) = x$. Then:

1. Start from an *initial guess* x_0.
2. For $k = 0, 1, 2, \ldots$, set

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

   until x_{k+1} satisfies termination criteria.

Fixed Point Theorem



Theorem: Fixed Point.
If $g \in C[a,b]$ and $a \leq g(x) \leq b$ for all $x \in [a,b]$, then there is a fixed point x^* in the interval $[a,b]$.

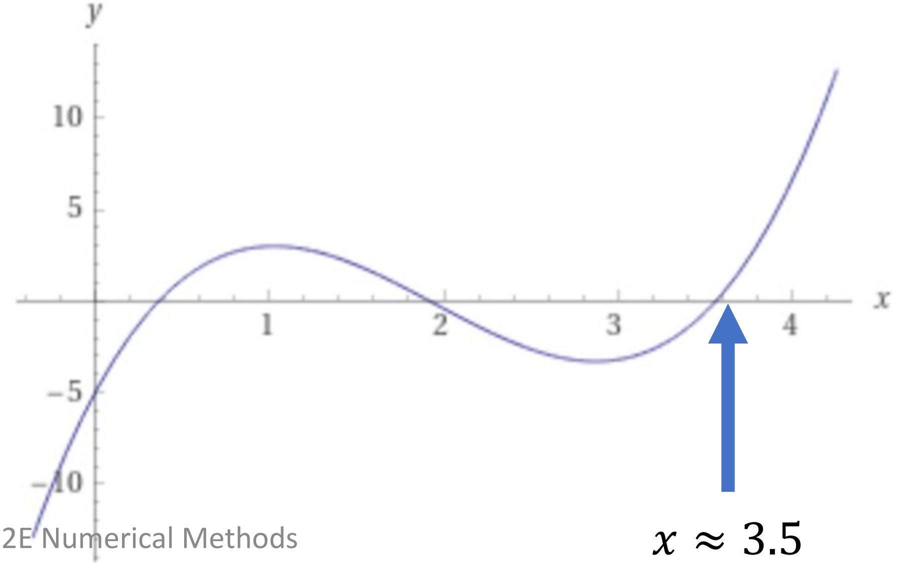
If, in addition, the derivative g' exists and there is a constant $\rho < 1$ such that the derivative satisfies

$$|g'(x)| \leq \rho \quad \forall\, x \in (a,b),$$

then the fixed point x^* is unique in this interval.

Fixed Point Iteration Example

- Find the highest real root of $\boldsymbol{f(x) = 2x^3 - 11.7x^2 + 17.7x - 5}$
- Compute the approximate percent relative errors for your solutions.
- Number of Iterations $\boldsymbol{(n) : 3}$
- Initial Guess $\boldsymbol{(x_0) : 3}$
- Graphical Representation:



$x \approx 3.5$

Fixed Point Iteration Solution

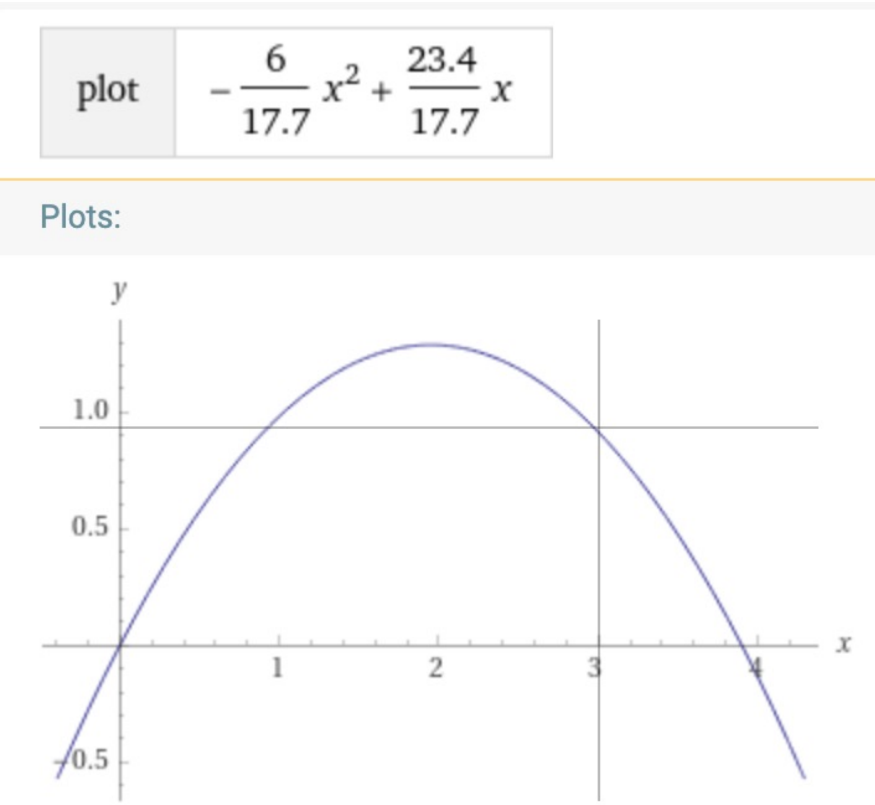
- Let’s find appropriate g(x) function:
- $f(x) = 2x^3 - 11.7x^2 + 17.7x - 5$
- $0 = 2x^3 - 11.7x^2 + 17.7x - 5$
- $-17.7x = 2x^3 - 11.7x^2 - 5$
- $17.7x = -2x^3 + 11.7x^2 + 5$
- $x = \frac{-2x^3+11.7x^2+5}{17.7}$

$$\boxed{\mathrm{g}(x) = x = \frac{-2x^3+11.7x^2+5}{17.7}}$$

Uniqueness of Fixed Point

$$g'(x) = \frac{-6}{17.7}x^2+\frac{11.7 \cdot 2x}{17.7} < 1 \Leftrightarrow -6x^2+23.4x-17.7 < 0 \Leftrightarrow x < 1.0267 \ \vee \ x > 2.8733$$

- We need to check that the derivative is bounded by some number less than 1 in the interval [a,b] that we are searching for the fixed point.
- Set the derivative to equal less than 1.
- Solve for x.
- Here, if we take our interval to be 3 and 4, there is a fixed point that is unique. Since x is less than 0 after x=2.8733.
- However, if we take that interval to be 0 and 4, then we cannot bound the derivative function. Therefore the fixed point is not unique. This can be verified by the previous slide.



plot $-\frac{6}{17.7}x^2+\frac{23.4}{17.7}x$

Plots:

Fixed Point Iteration Solution

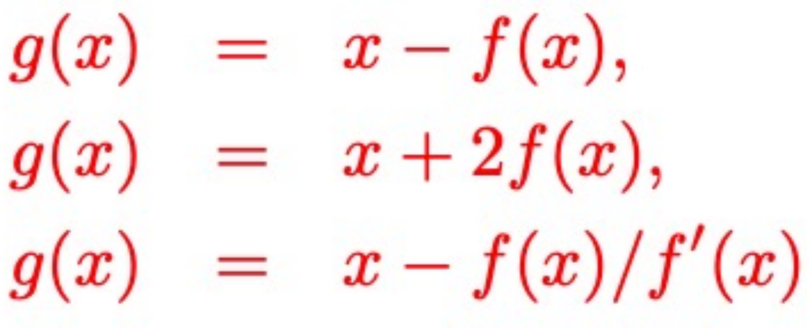
- $f(x) = 2x^3 - 11.7x^2 + 17.7x - 5$
- $\mathrm{g}(x) = x = \frac{-2x^3+11.7x^2+5}{17.7}$
- $x_0 = 3$
- I1: $x_1 = g(x_0) = g(3) = \frac{-2(3)^3+11.7(3)^2+5}{17.7} = 3.18079$ $\quad \varepsilon = \frac{3.18079 - 3}{3.18079} = 0.0568$
- I2: $x_2 = g(x_1) = \frac{-2(3.18079)^3+11.7(3.18079)^2+5}{17.7} = 3.33396$ $\quad \varepsilon = \frac{3.33396 - 3.18079}{3.33396} = 0.0459$
- I2: $x_3 = g(x_2) = \frac{-2(3.33396)^3+11.7(3.33396)^2+5}{17.7} = 3.44255$ $\quad \varepsilon = \frac{3.44255 - 3.33396}{3.44255} = 0.0315$

$$\boxed{f(x) = 0 \implies x \approx 3.44255}$$

CHP 3 Review Questions:

In what way is the fixed point iteration a family of methods, rather than just one method like bisection or secant?

- Given $f(x) = 0$, fixed point iteration requires that we rearrange to obtain $g(x) = x$ and iterate $x_{k+1} = g(x_k), k = 0, 1, 2, \ldots$
- There are many ways to obtain $g(x)$

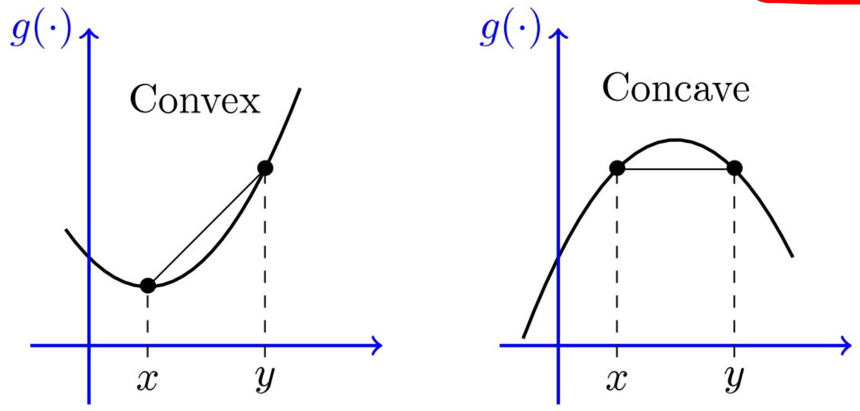


$$\begin{aligned} g(x) &= x - f(x), \\ g(x) &= x + 2f(x), \\ g(x) &= x - f(x)/f'(x) \end{aligned}$$

CHP 3 Review Questions:

State what a convex function is, and explain what happens if an objective function is convex.

- In mathematics, a real-valued function defined on an n-dimensional interval is called convex if the line segment between any two points on the graph of the function lies above the graph between the two points.
- **if** the **objective function** is strictly **convex**, then the problem has at most one optimal point.



CHP 4

Basic Concepts

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 &= b_1,\\
a_{21}x_1 + a_{22}x_2 &= b_2.
\end{aligned}
$$

$$\Downarrow$$

$$\mathbf{a}_1 x_1 + \mathbf{a}_2 x_2 + \cdots + \mathbf{a}_n x_n = \mathbf{b},$$

$$\Downarrow$$

$$\begin{pmatrix} \vdots \\ \vdots \\ \mathbf{a}_1 \\ \vdots \\ \vdots \end{pmatrix} x_1 + \begin{pmatrix} \vdots \\ \vdots \\ \mathbf{a}_2 \\ \vdots \\ \vdots \end{pmatrix} x_2 + \cdots + \begin{pmatrix} \vdots \\ \vdots \\ \mathbf{a}_n \\ \vdots \\ \vdots \end{pmatrix} x_n = \begin{pmatrix} \vdots \\ \vdots \\ \mathbf{b} \\ \vdots \\ \vdots \end{pmatrix} \quad \text{and} \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix}$$

<u>Equivalent Statements</u>

A is nonsingular.

$\det(A) \neq 0$.

The columns of A are linearly independent.

The rows of A are linearly independent.

There exists a matrix which we denote by A^{-1}, which satisfies $A^{-1}A = I = AA^{-1}$.

$\text{range}(A) = \mathcal{R}^n$.

$\text{null}(A) = \{0\}$.

Eigenvalues and Eigenvectors

- **Eigenvectors** are special vectors that when a linear transformation is applied, they are only scaled by a constant value. This constant value is the corresponding **eigenvalue**.

- We can find eigenvalues by solving the characteristic polynomial:

$$\det(\lambda I - A) = 0$$

Symmetric Positive Definite Matrices

- Symmetric: $A^T = A$
- Positive Definite: $\mathbf{x}^T A\mathbf{x} > 0 \quad \forall\, \mathbf{x} \neq \mathbf{0}$
- Extension of positive scalars to n dimensions.

- RMK: A symmetric matrix is positive definite iff all of its eigenvalues are positive

Eigenvalue and Sym. Pos. Def. Matrix Example

- Determine if the following matrix is symmetric positive definite:
- $A = \begin{matrix} 1 & 2 \\ 2 & 1 \end{matrix}$
- Symmetric ✓
- $\det(\lambda I - A) = \begin{vmatrix} \lambda - 1 & -2 \\ -2 & \lambda - 1 \end{vmatrix} = (\lambda - 1)(\lambda - 1) - 4$
- $(\lambda - 1)(\lambda - 1) - 4 = 0$
- $\lambda^2 - 2\lambda + 1 - 4 = 0$
- $\lambda^2 - 2\lambda - 3 = 0$
- $(\lambda - 3)(\lambda + 1) \ = 0$
- $\boldsymbol{\lambda_1 = -1}$ **&** $\boldsymbol{\lambda_2 = 3}$

Symmetric but **not** positive definite!

Eigenvalue and Sym. Pos. Def. Matrix Example

- Determine if the following matrix is symmetric positive definite:
- $A = \begin{matrix} 5 & 4 \\ 4 & 5 \end{matrix}$
- Symmetric ✓
- $\det(\lambda I - A) = \begin{vmatrix} \lambda - 5 & -4 \\ -4 & \lambda - 5 \end{vmatrix} = (\lambda - 5)(\lambda - 5) - 16$
- $(\lambda - 5)(\lambda - 5) - 16 = 0$
- $\lambda^2 - 10\lambda + 25 - 16 = 0$
- $\lambda^2 - 10\lambda + 9 = 0$
- $(\lambda - 9)(\lambda - 1) \ = 0$
- $\boldsymbol{\lambda_1 = 9}$ & $\boldsymbol{\lambda_2 = 1}$

Symmetric positive definite!

Eigenvalue and Sym. Pos. Def. Matrix Example

- Determine if the following matrix is symmetric positive definite:
- $A = \begin{matrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{matrix}$
- Symmetric ✓
- $\det(A - \lambda I) = \begin{vmatrix} 1-\lambda & 0 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 0 & 3-\lambda \end{vmatrix} = (1-\lambda)(2-\lambda)(3-\lambda)$
- $\boldsymbol{\lambda_1 = 1}$ & $\boldsymbol{\lambda_2 = 2}$ & $\boldsymbol{\lambda_3 = 3}$

Symmetric positive definite!

CHP 4 Review Questions:

What is a singular matrix?

- When the determinant of a square matrix *A* equals 0 we say that *A* is **singular, or noninvertible**.
- Equivalently, *A* is nonsingular if and only if its columns (or rows) are **linearly independent**. Then, A is said to be invertible.

CHP 4 Review Questions:

Suppose $Ax = b \neq 0$ is a linear system and A is a square, singular matrix. How many solutions is it possible for the system to have?

- If *A* is *n* × *n* and singular, then there are vectors in R*n* which do not belong to range(*A*), and there are nonzero vectors which belong to *A*'s nullspace, null(*A*).
- So, in this case, for a particular given right-hand-side vector **b** the equation *A***x** = **b** will not have a solution if **b** ∉ range(*A*).
- On the other hand, if *A* is singular and **b** ∈ range(*A*), then there are infinitely many solutions.
- If A does not have a pivot in every row, that does not mean that Ax = b does not have a solution for some given vector b. It just means that there are some vectors b for which Ax = b does not have a solution.

CHP 4 Review Questions:

What is the spectrum of a matrix?

- **spectral** radius of a square **matrix** is the largest absolute value of its eigenvalues
- Here the product of *A* and the **eigenvector x** equals the product of the scalar **eigenvalue** λ and the same vector. We require $\mathbf{x} \neq \mathbf{0}$ in order not to say something trivial.
- Together, (λ, **x**) is an **eigenpair**, and the set of eigenvalues forms the **spectrum** of *A*.

CHP 4 Review Questions:

Define algebraic multiplicity and geometric multiplicity of eigenvalues.

- If there are *k* roots that are equal to each other, call their value λ*j*; then we say that λ*j* is an eigenvalue with **algebraic multiplicity** *k*.
- **geometric multiplicity** is defined as the dimension of the space that its associated eigenvectors span; in other words, it is the number of linearly independent eigenvectors associated with this eigenvalue.

CHP 4 Review Questions:

What is a diagonalizable matrix?

- A is diagonalizable if it can be written as

$$A = X \Lambda X^{-1}$$

- Where Λ is the diagonal matrix of eigenvalues and X is a matrix composed of the eigenvectors of A.

CHP 4 Review Questions:

How is the norm of a vector related to the absolute value of a scalar?

- They both show magnitude without direction.

CHP 4 Review Questions:

Is it true that every symmetric positive definite matrix is necessarily nonsingular?

- The determinant of a **positive definite matrix** is always **positive**, so a **positive definite matrix** is always **nonsingular**.

CHP 4 Review Questions:

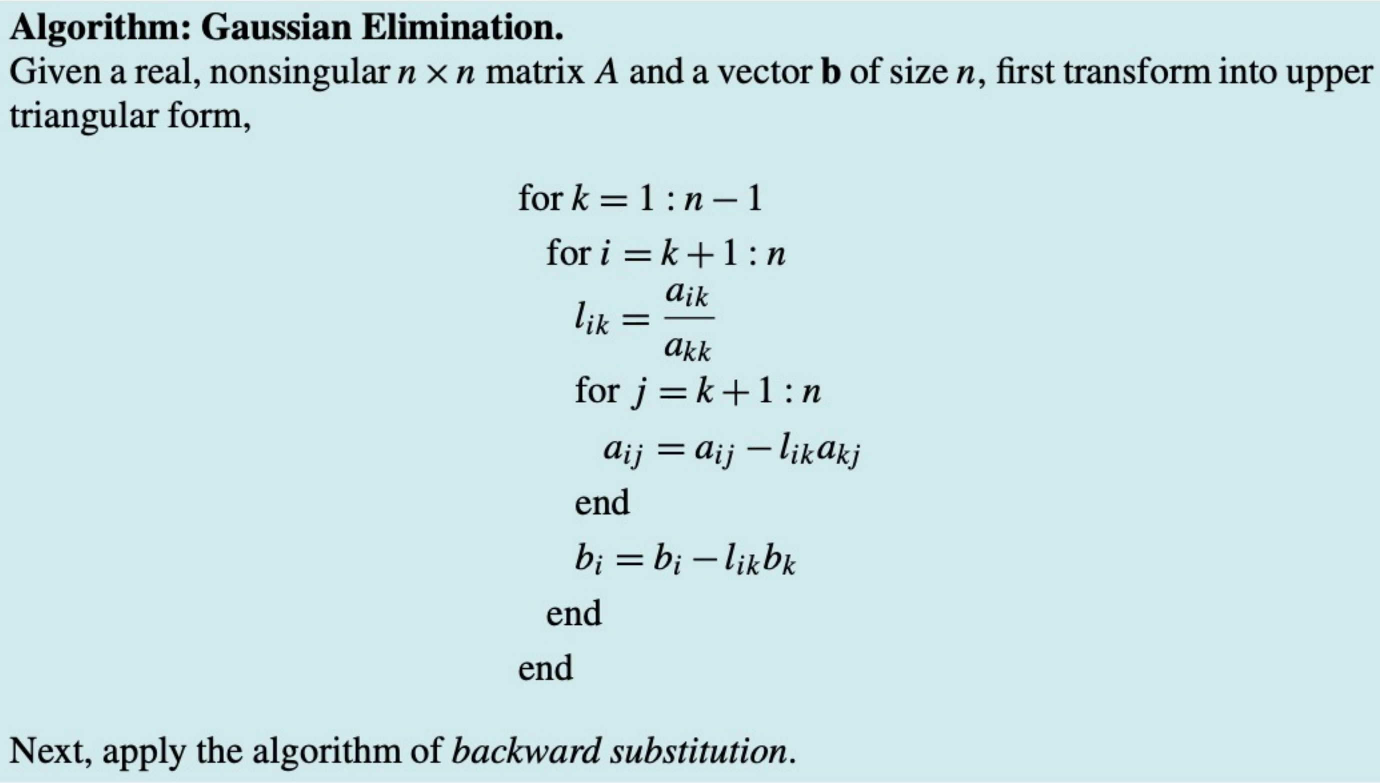
Give an example of a 3 × 3 orthogonal matrix not equal to the identity.

- To **check if** a given **matrix is orthogonal**, first find the transpose of that **matrix**. Then, multiply the given **matrix** with the transpose. Now, **if** the product is an identity **matrix**, the given **matrix is orthogonal**, otherwise, not.

$$\begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} \quad \text{since} \quad \begin{matrix} 0 & 1 & 0^{T} \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} = \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} \quad \text{and} \quad \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} * \begin{matrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{matrix} = \begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix} = I_3$$

CHP 5

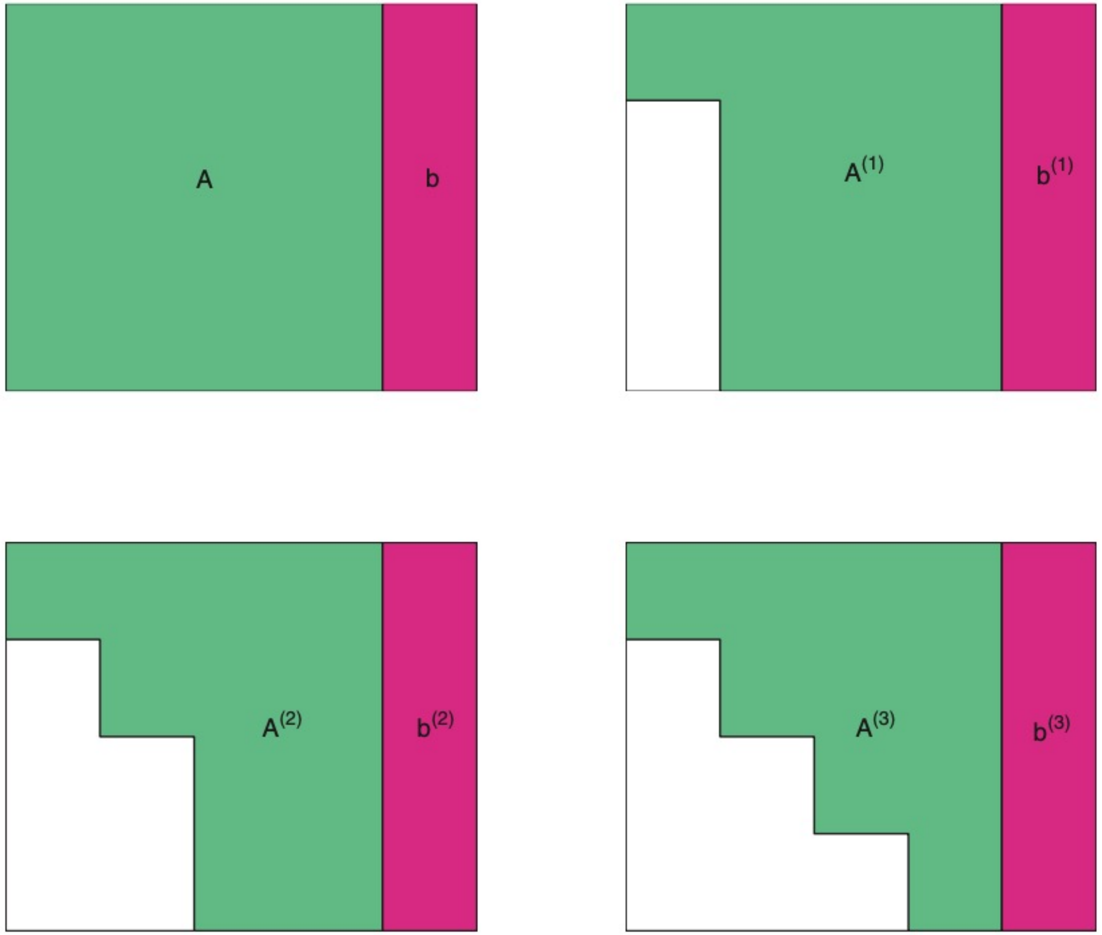
Gaussian Elimination



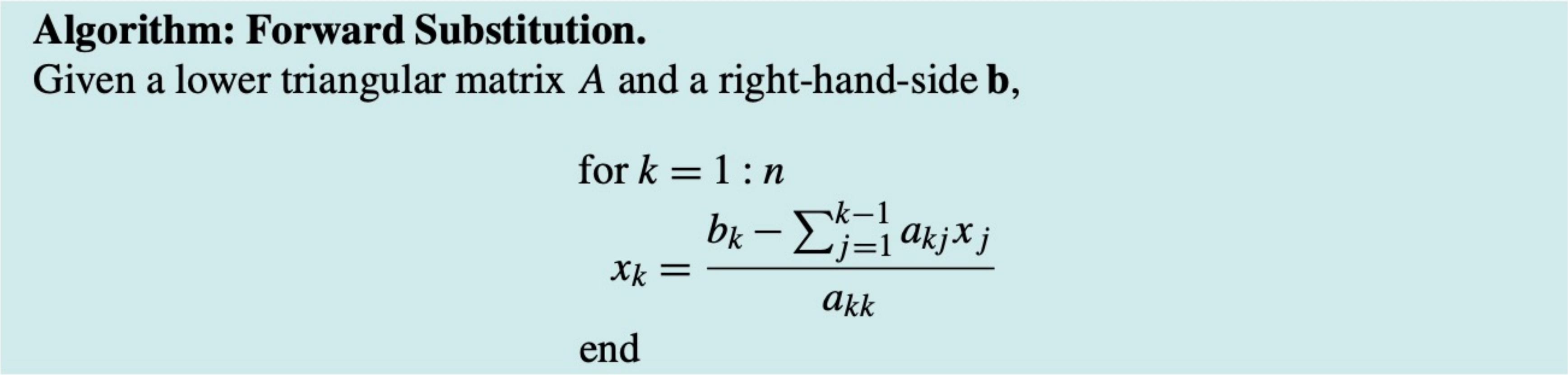
Algorithm: Gaussian Elimination.
Given a real, nonsingular $n \times n$ matrix A and a vector $\mathbf{b}$ of size n, first transform into upper triangular form,

```
for $k = 1 : n-1$
  for $i = k+1 : n$
    $l_{ik} = \frac{a_{ik}}{a_{kk}}$
    for $j = k+1 : n$
      $a_{ij} = a_{ij} - l_{ik} a_{kj}$
    end
    $b_i = b_i - l_{ik} b_k$
  end
end
```

Next, apply the algorithm of *backward substitution*.

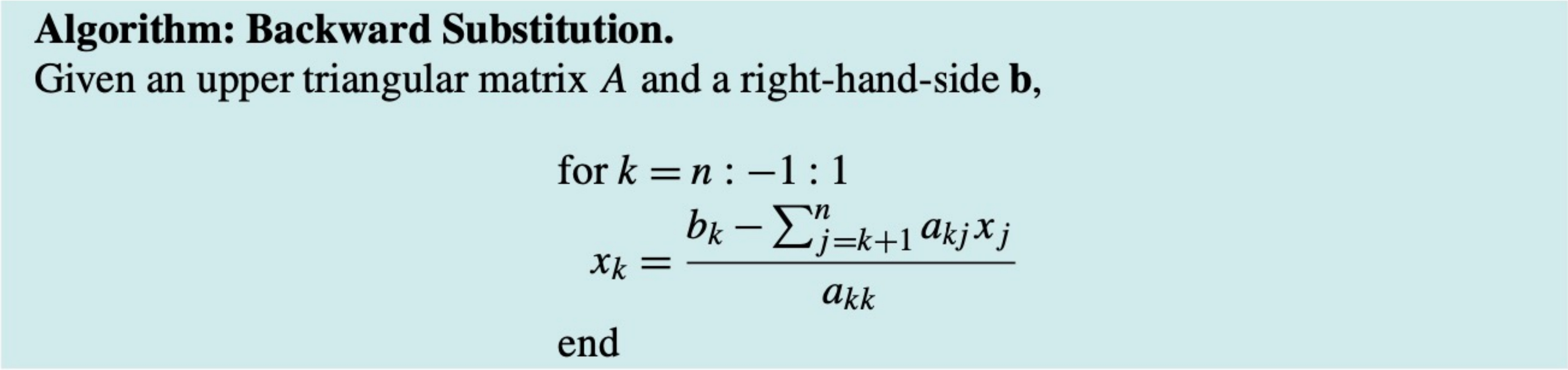


Forward/Backward Substitution



Algorithm: Forward Substitution.
Given a lower triangular matrix A and a right-hand-side $\mathbf{b}$,

```
for $k = 1 : n$
    $x_k = \dfrac{b_k - \sum_{j=1}^{k-1} a_{kj} x_j}{a_{kk}}$
end
```



Algorithm: Backward Substitution.
Given an upper triangular matrix A and a right-hand-side $\mathbf{b}$,

```
for $k = n : -1 : 1$
    $x_k = \dfrac{b_k - \sum_{j=k+1}^{n} a_{kj} x_j}{a_{kk}}$
end
```

GE with BW Substitution Example

- Use Gaussian Elimination and Backward Substitution to solve:
- $8x_1 + 2x_2 - 2x_3 = -2$
- $10x_1 + 2x_2 + 4x_3 = 4$
- $12x_1 + 2x_2 + 2x_3 = 6$
- Augmented Matrix:

$$\begin{vmatrix} 8 & 2 & -2 & -2 \\ 10 & 2 & 4 & 4 \\ 12 & 2 & 2 & 6 \end{vmatrix}$$

Gaussian Elimination

$$\begin{vmatrix} 8 & 2 & -2 & -2 \\ 10 & 2 & 4 & 4 \\ 12 & 2 & 2 & 6 \end{vmatrix} \xrightarrow{(-10/8)R1 + R2} \begin{vmatrix} 8 & 2 & -2 & -2 \\ 0 & -0.5 & 6.5 & 6.5 \\ 12 & 2 & 2 & 6 \end{vmatrix} \xrightarrow{(-12/8)R1 + R2} \begin{vmatrix} 8 & 2 & -2 & -2 \\ 0 & -0.5 & 6.5 & 6.5 \\ 0 & -1 & 5 & 9 \end{vmatrix}$$

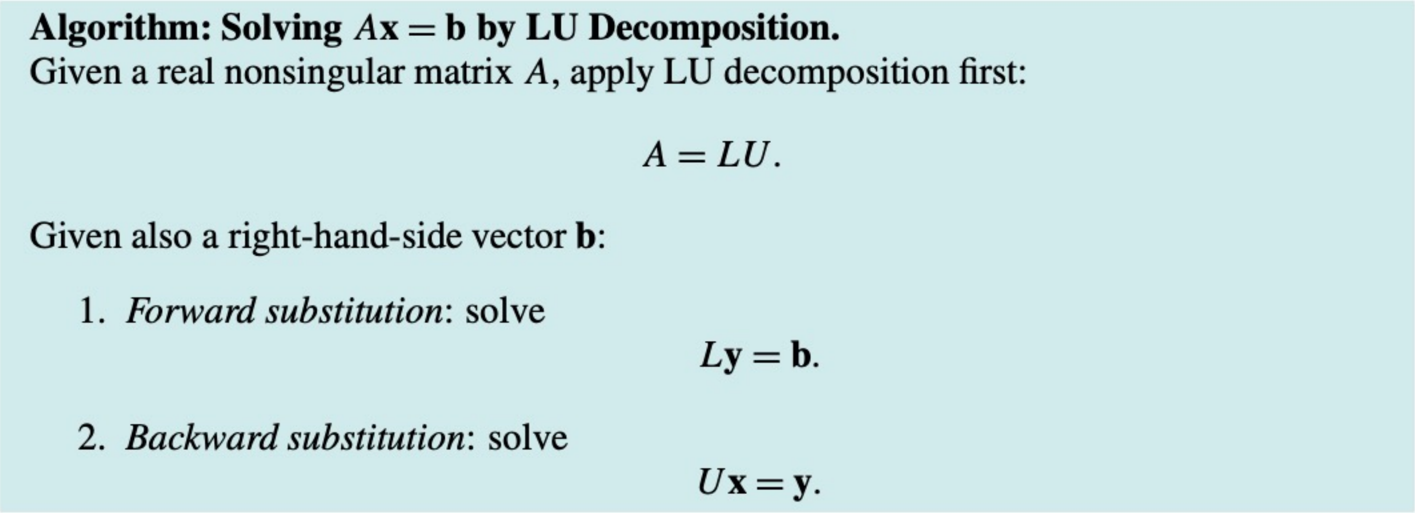
$$\begin{vmatrix} 8 & 2 & -2 & -2 \\ 0 & -0.5 & 6.5 & 6.5 \\ 0 & -1 & 5 & 9 \end{vmatrix} \xrightarrow{(-1/-0.5)R2 + R3} \begin{vmatrix} 8 & 2 & -2 & -2 \\ 0 & -0.5 & 6.5 & 6.5 \\ 0 & 0 & -8 & -4 \end{vmatrix}$$

Backward Substitution

$$\begin{vmatrix} 8 & 2 & -2 & -2 \\ 0 & -0.5 & 6.5 & 6.5 \\ 0 & 0 & -8 & -4 \end{vmatrix}$$

- $-8x_3 = -4 \ \rightarrow x_3 = \mathbf{0.5}$
- $-0.5x_2 + 6.5x_3 = 6.5 \ \rightarrow x_2 = \mathbf{-6.5}$
- $8x_1 + 2x_2 - 2x_3 = -2 \ \rightarrow x_1 = \mathbf{1.5}$

LU Decomposition

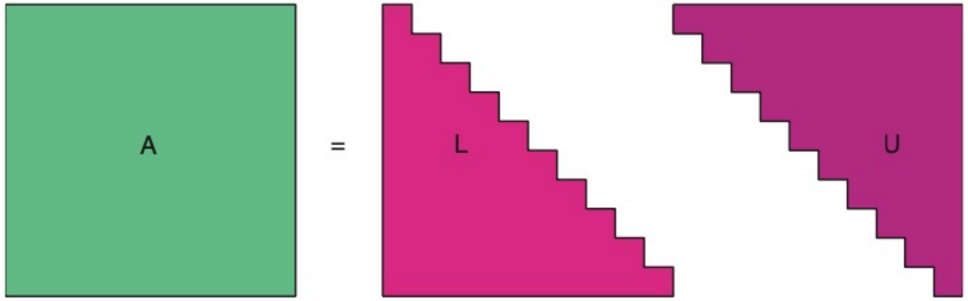


Algorithm: Solving $A\mathbf{x} = \mathbf{b}$ by LU Decomposition.
Given a real nonsingular matrix A, apply LU decomposition first:

$$A = LU.$$

Given also a right-hand-side vector $\mathbf{b}$:

1. *Forward substitution*: solve
$$L\mathbf{y} = \mathbf{b}.$$
2. *Backward substitution*: solve
$$U\mathbf{x} = \mathbf{y}.$$



LU Decomposition Method

- Alternative Method:
- **Step 1.** Write down the identity matrix of the same size on the left and the initial matrix on the right.
- **Step 2.** Apply Gaussian Elimination to the right hand side matrix. Use the left hand side matrix to remember your actions (If you multiply the right hand side by (2/8), then mark down the opposite sign number (-2/8) on the left hand side matrix in the same position where a term was canceled out on the right hand side matrix.)
- **Step 3.** Repeat until the matrix on left is lower triangular and matrix on the right is upper triangular. These are the L and U matrices respectively.

- THE MATRIX ON THE LEFT IS FOR MEMORIZATION OF ACTIONS. IT IS NOT A RESULT OF ELEMENTARY ROW OPERATIONS!!

* RMK: Please make sure you're also familiar with the exact method shown in class (i.e. computing and combining M matrices).

LU Decomposition Example

Compute the LU decomposition of the following matrix:

$$\begin{vmatrix} 8 & 4 & -1 \\ -2 & 5 & 1 \\ 2 & -1 & 6 \end{vmatrix}$$

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 8 & 4 & -1 \\ -2 & 5 & 1 \\ 2 & -1 & 6 \end{vmatrix}$$

$$\Big\downarrow \qquad \text{(2/8)R1 + R2} \qquad \Big\downarrow$$

$$\begin{vmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 2 & -1 & 6 \end{vmatrix}$$

LU Decomposition Example

$$\begin{vmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 2 & -1 & 6 \end{vmatrix}$$

$$\Big\downarrow \quad (-2/8)R1 + R3 \quad \Big\downarrow$$

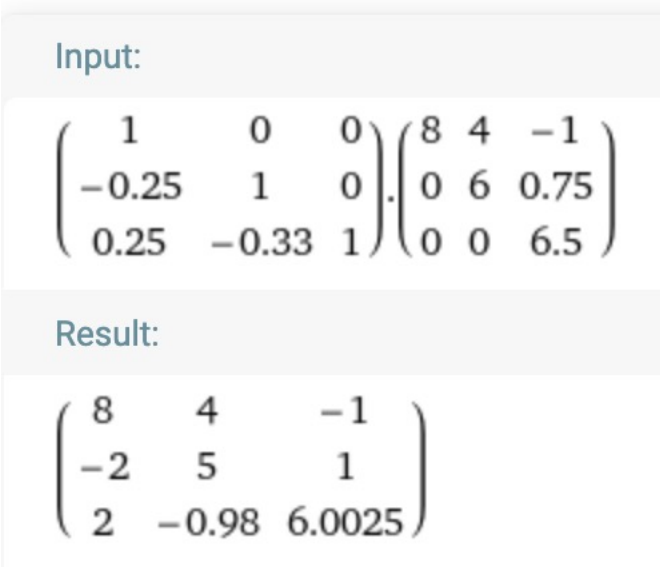
$$\begin{vmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 0 & -2 & 6.25 \end{vmatrix}$$

$$\longrightarrow \quad \begin{vmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & 0 & 1 \end{vmatrix} \quad \begin{vmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 0 & -2 & 6.25 \end{vmatrix}$$

$$\Big\downarrow \quad (2/6)R2 + R3 \quad \Big\downarrow$$

$$\boxed{\underbrace{\begin{vmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & -0.33 & 1 \end{vmatrix}}_{L} \quad \underbrace{\begin{vmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 0 & 0 & 6.5 \end{vmatrix}}_{U}}$$

Let’s Check



Input:

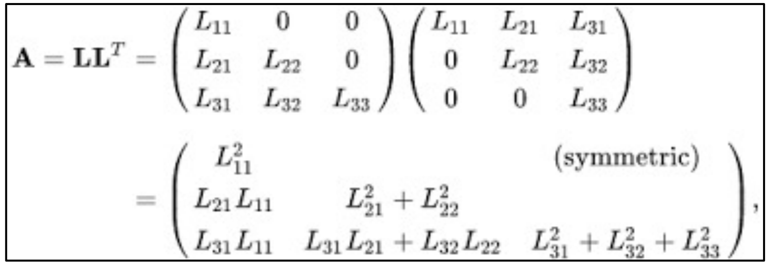
$$\begin{pmatrix} 1 & 0 & 0 \\ -0.25 & 1 & 0 \\ 0.25 & -0.33 & 1 \end{pmatrix} . \begin{pmatrix} 8 & 4 & -1 \\ 0 & 6 & 0.75 \\ 0 & 0 & 6.5 \end{pmatrix}$$

Result:

$$\begin{pmatrix} 8 & 4 & -1 \\ -2 & 5 & 1 \\ 2 & -0.98 & 6.0025 \end{pmatrix}$$

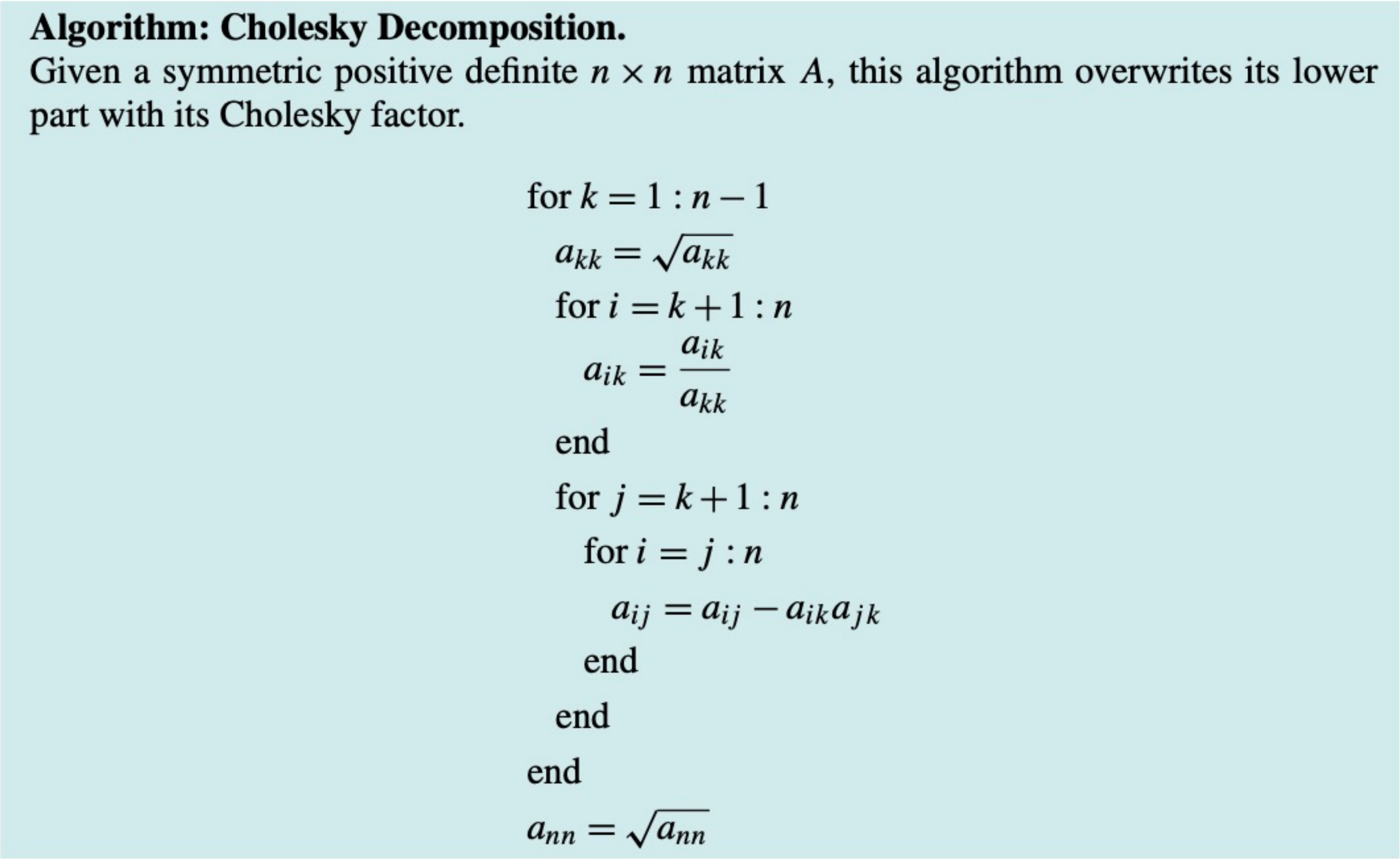
Another example employing the same method can be found at:
https://www.youtube.com/watch?v=BFYFkn-eOQk

Cholesky Decomposition



$$\mathbf{A} = \mathbf{L}\mathbf{L}^T = \begin{pmatrix} L_{11} & 0 & 0 \\ L_{21} & L_{22} & 0 \\ L_{31} & L_{32} & L_{33} \end{pmatrix} \begin{pmatrix} L_{11} & L_{21} & L_{31} \\ 0 & L_{22} & L_{32} \\ 0 & 0 & L_{33} \end{pmatrix}$$

$$= \begin{pmatrix} L_{11}^2 & & \text{(symmetric)} \\ L_{21}L_{11} & L_{21}^2 + L_{22}^2 & \\ L_{31}L_{11} & L_{31}L_{21} + L_{32}L_{22} & L_{31}^2 + L_{32}^2 + L_{33}^2 \end{pmatrix},$$



Algorithm: Cholesky Decomposition.

Given a symmetric positive definite $n \times n$ matrix A, this algorithm overwrites its lower part with its Cholesky factor.

```
for $k = 1 : n-1$
  $a_{kk} = \sqrt{a_{kk}}$
  for $i = k+1 : n$
    $a_{ik} = \frac{a_{ik}}{a_{kk}}$
  end
  for $j = k+1 : n$
    for $i = j : n$
      $a_{ij} = a_{ij} - a_{ik}a_{jk}$
    end
  end
end
$a_{nn} = \sqrt{a_{nn}}$
```

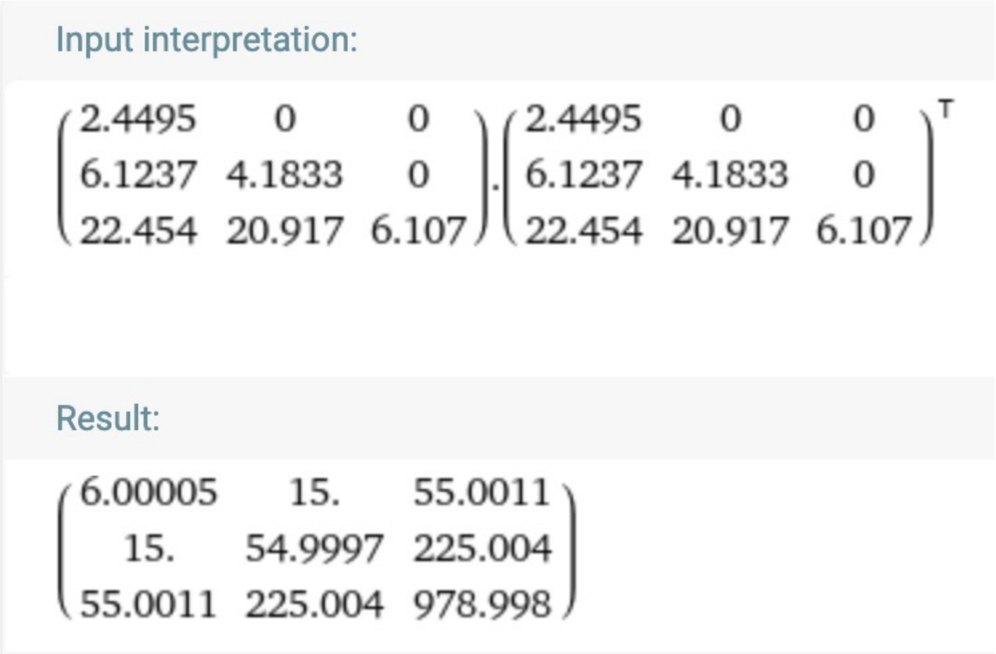
Cholesky Decomposition Example

- Apply Cholesky decomposition to the symmetric matrix: **A =** $\begin{vmatrix} 6 & 15 & 55 \\ 15 & 55 & 225 \\ 55 & 225 & 979 \end{vmatrix}$
- $l_{11} = \sqrt{a_{11}} = \sqrt{6} = 2.4495$
- $l_{21} = \frac{a_{21}}{l11} = \frac{15}{2.4495} = 6.1237$
- $l_{22} = \sqrt{a_{22} - {l_{21}}^2} = \sqrt{55 - 6.1237^2} = 4.1833$
- $l_{31} = \frac{a_{31}}{l_{11}} = \frac{55}{2.4495} = 22.454$
- $l_{32} = \frac{a_{32} - l_{21} l_{31}}{l_{22}} = \frac{225 - (6.1237)(22.454)}{4.1833} = 20.917$
- $l_{33} = \sqrt{a_{33} - {l_{31}}^2 - {l_{32}}^2} = \sqrt{979 - 22.454^2 - 20.917^2} = 6.107$

L = $\begin{vmatrix} 2.4495 & 0 & 0 \\ 6.1237 & 4.1833 & 0 \\ 22.454 & 20.917 & 6.107 \end{vmatrix}$

Let’s Check

$A = LL^T$



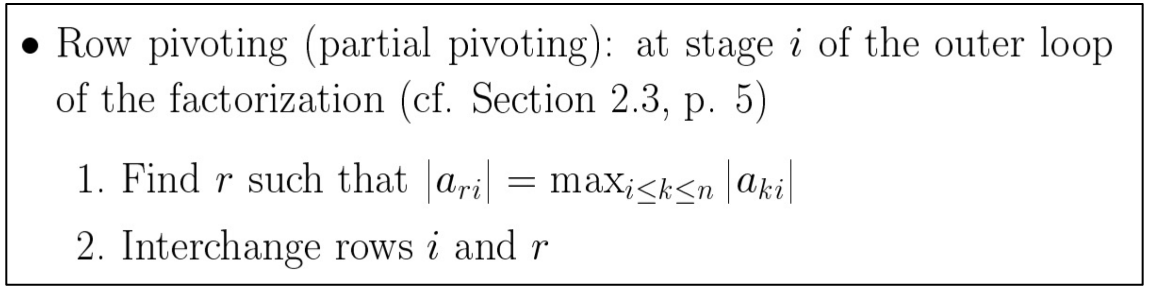
Input interpretation:

$$\begin{pmatrix} 2.4495 & 0 & 0 \\ 6.1237 & 4.1833 & 0 \\ 22.454 & 20.917 & 6.107 \end{pmatrix} . \begin{pmatrix} 2.4495 & 0 & 0 \\ 6.1237 & 4.1833 & 0 \\ 22.454 & 20.917 & 6.107 \end{pmatrix}^T$$

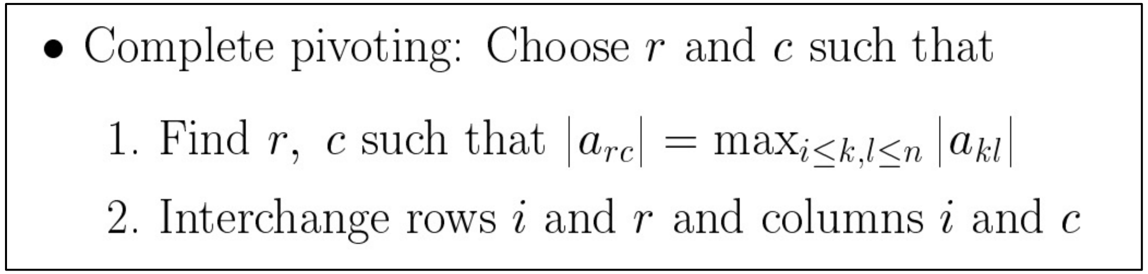
Result:

$$\begin{pmatrix} 6.00005 & 15. & 55.0011 \\ 15. & 54.9997 & 225.004 \\ 55.0011 & 225.004 & 978.998 \end{pmatrix}$$

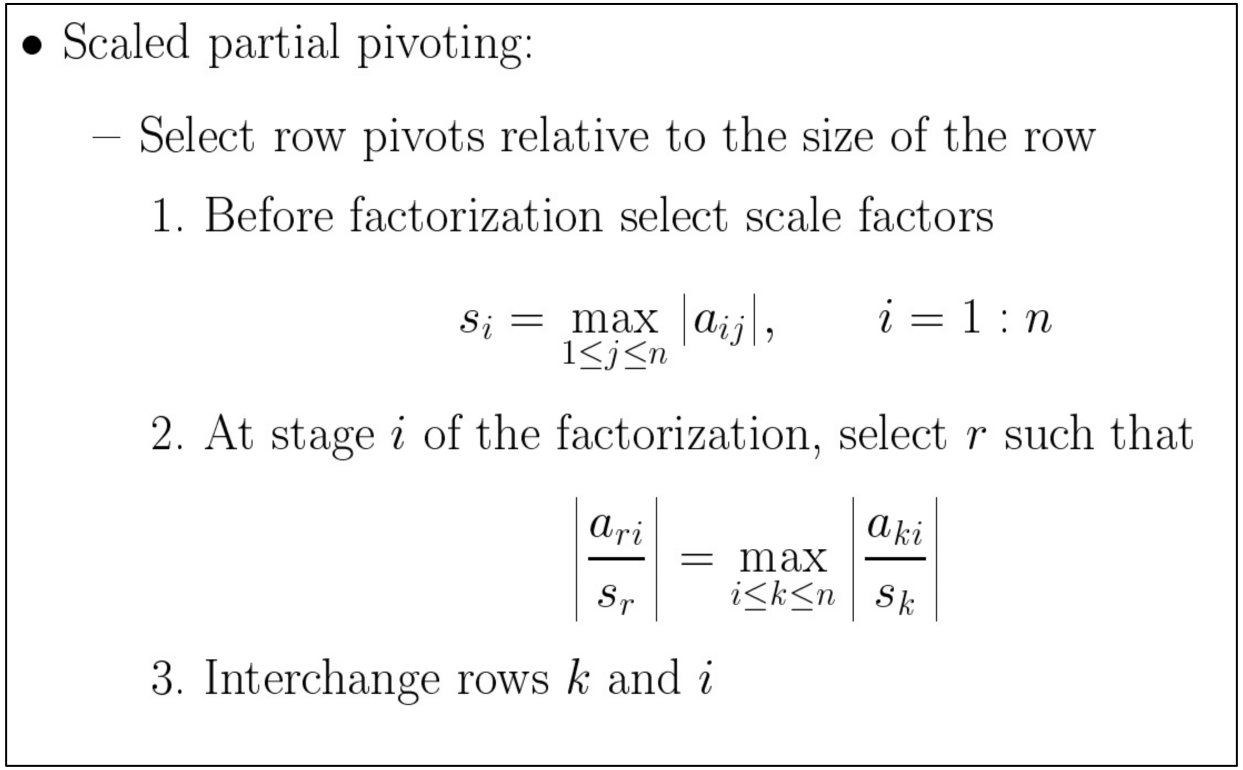
Pivoting



- Row pivoting (partial pivoting): at stage i of the outer loop of the factorization (cf. Section 2.3, p. 5)
  1. Find r such that $|a_{ri}| = \max_{i \le k \le n} |a_{ki}|$
  2. Interchange rows i and r



- Complete pivoting: Choose r and c such that
  1. Find $r,\ c$ such that $|a_{rc}| = \max_{i \le k, l \le n} |a_{kl}|$
  2. Interchange rows i and r and columns i and c



- Scaled partial pivoting:
  – Select row pivots relative to the size of the row
  1. Before factorization select scale factors

$$s_i = \max_{1 \le j \le n} |a_{ij}|, \qquad i = 1 : n$$

  2. At stage i of the factorization, select r such that

$$\left|\frac{a_{ri}}{s_r}\right| = \max_{i \le k \le n} \left|\frac{a_{ki}}{s_k}\right|$$

  3. Interchange rows k and i

Partial Pivoting Example

$$\left|\begin{array}{cccc} 8 & 2 & -2 & -2 \\ 10 & 2 & 4 & 4 \\ 12 & 2 & 2 & 6 \end{array}\right| \xrightarrow{\text{Swap R1 R3}} \left|\begin{array}{cccc} 12 & 2 & 2 & 6 \\ 10 & 2 & 4 & 4 \\ 8 & 2 & -2 & -2 \end{array}\right| \xrightarrow[\text{(-8/12)R1+R3}]{\text{(-10/12)R1+R2}} \left|\begin{array}{cccc} 12 & 2 & 2 & 6 \\ 0 & 1/3 & 7/3 & -1 \\ 0 & 2/3 & -10/3 & -6 \end{array}\right|$$

$$\xrightarrow{\text{Swap R2 R3}} \left|\begin{array}{cccc} 12 & 2 & 2 & 6 \\ 0 & 2/3 & -10/3 & -6 \\ 0 & 1/3 & 7/3 & -1 \end{array}\right| \xrightarrow{\text{(-1/2)R2+R3}} \left|\begin{array}{cccc} 12 & 2 & 2 & 6 \\ 0 & 2/3 & -10/3 & -6 \\ 0 & 0 & 4 & 2 \end{array}\right|$$

$$\begin{aligned} 4x_3 = 2 &\Rightarrow x_3 = \boxed{0.5} \\ \frac{2}{3}x_2 - \frac{10}{3}x_3 = -6 &\Rightarrow x_2 = \boxed{-6.5} \\ 12x_1 + 2x_2 + 2x_3 = 6 &\Rightarrow x_1 = \boxed{1.5} \end{aligned}$$

Scaled Partial Pivoting Example

Starting Values

L	1	2	3

S	2	1	3

$$\left|\begin{array}{ccc|c|} 1 & -1 & 2 & 2 \\ 1 & -1 & 1 & 1 \\ 2 & 3 & -1 & 4 \end{array}\right.$$

Check scaling factors:

$$\{\frac{1}{2},\frac{\mathbf{1}}{\mathbf{1}},\frac{2}{3}\}$$

EQN 2 becomes pivot

Update:

L	2	1	3

- Row reduce using new pivot.
- Then, repeat for remaining sub matrix.

$$\left|\begin{array}{ccc|c|} 1 & -1 & 2 & 2 \\ 1 & -1 & 1 & 1 \\ 2 & 3 & -1 & 4 \end{array}\right. \xrightarrow[(-2/1)R2+R3]{(-1/1)R2+R1} \left|\begin{array}{ccc|c|} 0 & 0 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 0 & 5 & -3 & 2 \end{array}\right|$$

$$\left|\begin{array}{ccc|c|} 0 & 0 & 1 & 1 \\ 1 & -1 & 1 & 1 \\ 0 & 5 & -3 & 2 \end{array}\right.$$

Check scaling factors:

$$\{\frac{0}{2},\frac{\mathbf{5}}{\mathbf{3}}\}$$

EQN 3 becomes pivot

Update:

L	2	3	1

No more row operations are necessary after this point. We can use backward substitution to solve.

$$\boxed{\begin{aligned} \boldsymbol{x_3} &= \mathbf{1} \\ \boldsymbol{x_2} &= \mathbf{1} \\ \boldsymbol{x_1} &= \mathbf{1} \end{aligned}}$$

Condition Number

$$\kappa(A) = \|A\| \|A^{-1}\|$$

- For a symmetric positive definite matrix:
- $\|A\|_2 = \lambda_1$
- $\|A^{-1}\|_2 = \frac{1}{\lambda_n}$
- $\kappa(A) = \|A\| \|A^{-1}\| = \frac{\lambda_1}{\lambda_n}$

Condition Number Example

- From previous example…
- $A = \begin{matrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{matrix}$
- $\boldsymbol{\lambda_1 = 3}, \boldsymbol{\lambda_2 = 2}, \& \boldsymbol{\lambda_3 = 1}$
- $\kappa(A) = \|A\|\|A^{-1}\| = \frac{3}{1} = 3$

CHP 5 Review Questions:

During the course of Gaussian elimination without pivoting a zero pivot has been encountered. Is the matrix singular?

- Not necessarily:

$$A = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

CHP 5 Review Questions:

State three disadvantages of computing the inverse of a matrix to solve a linear system rather than using the LU decomposition approach.

- Forming the **inverse** is wasteful in storage.
- **Computing the inverse** is computationally more expensive.
- May give rise to a more pronounced presence of roundoff errors.

CHP 5 Review Questions:

The complete pivoting strategy is numerically stable, whereas partial pivoting is not always stable. Why is the latter approach preferred in practice in spite of this?

- Complete pivoting is computationally more expensive. We do not need the best pivot; we only need to avoid bad pivots.

CHP 5 Review Questions:

The condition number is an important concept, but it is rarely computed exactly. Why?

- Computing **the condition number** requires that we **compute** the inverse, which is computationally expensive

CHP 6

Linear Least Squares

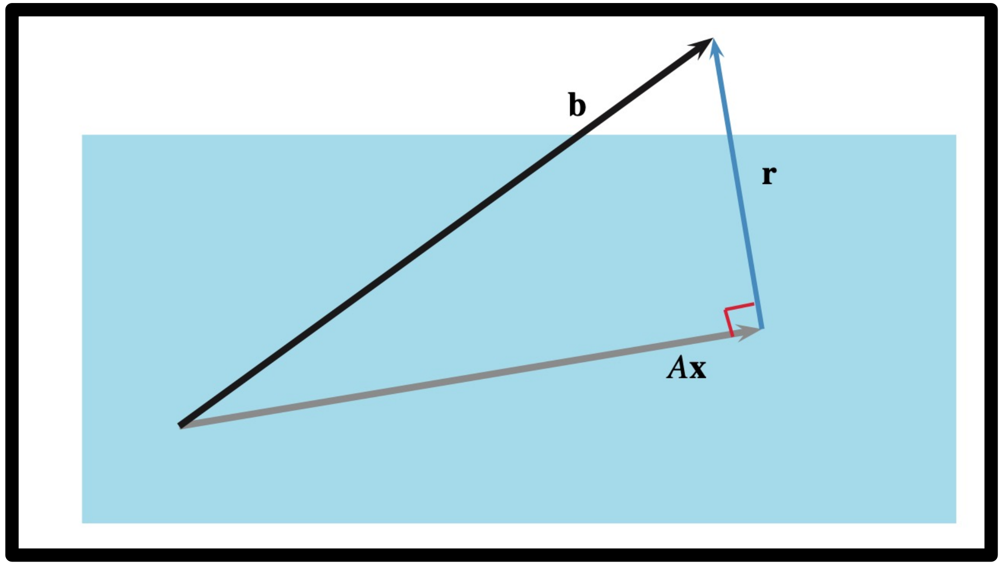
$$\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \; A = \begin{pmatrix} a_{1,1} & \dots & a_{1,n} \\ \vdots & & \vdots \\ \vdots & \dots & \vdots \\ a_{m,1} & \dots & a_{m,n} \end{pmatrix}, \; \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ \vdots \\ b_m \end{pmatrix}, \; \mathbf{r} = \begin{pmatrix} r_1 \\ \vdots \\ \vdots \\ r_m \end{pmatrix}$$

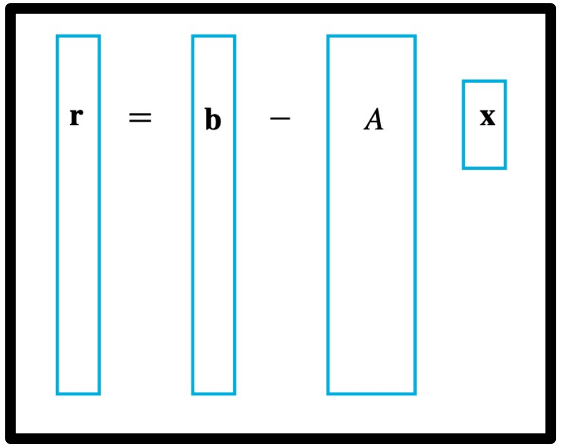
$\min_{\mathbf{x}} \|\mathbf{b} - A\mathbf{x}\|_2$ where $A \in \mathcal{R}^{m \times n}$, $\mathbf{x} \in \mathcal{R}^n$, $\mathbf{b} \in \mathcal{R}^m$, $m \geq n$

To simplify the math, rewrite:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{b} - A\mathbf{x}\|^2$$

$$\mathbf{r} = \mathbf{b} - A\mathbf{x}$$





Minimization of Objective Function

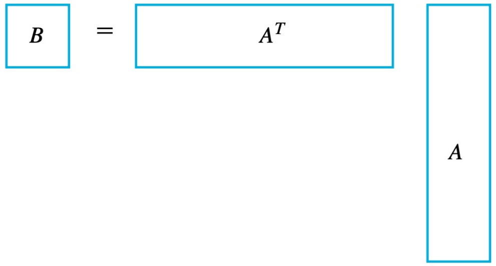
Our objective function is: $\min_{\mathbf{x}} \psi(\mathbf{x}), \quad \text{where} \;\; \psi(\mathbf{x}) = \frac{1}{2}\|\mathbf{r}\|^2 = \frac{1}{2}\sum_{i=1}^{m}\left(b_i - \sum_{j=1}^{n} a_{i,j}x_j\right)^2$

To solve, we take the derivative with respect to x and set it equal to 0

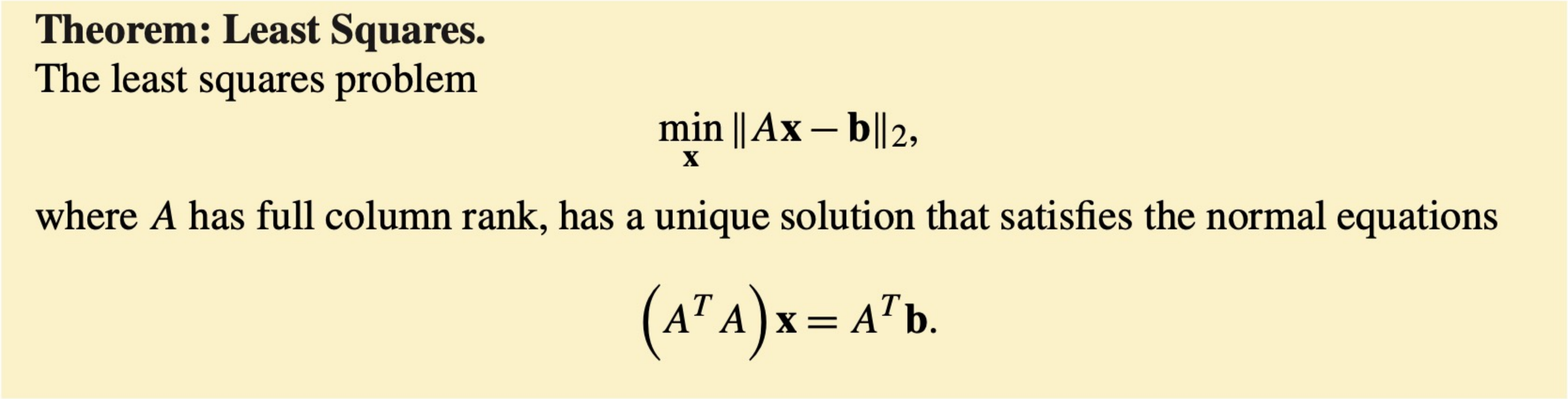
$$\frac{\partial}{\partial x_k}\psi(\mathbf{x}) = \sum_{i=1}^{m}\left[\left(b_i - \sum_{j=1}^{n} a_{i,j}x_j\right)(-a_{i,k})\right] = 0$$

In matrix vector form: $\underbrace{A^T A}_{\mathbf{B}}\mathbf{x} = A^T\mathbf{b}$

Symmetric
Positive Definite



Least Squares Theorem



Theorem: Least Squares.
The least squares problem

$$\min_{\mathbf{x}} \|A\mathbf{x} - \mathbf{b}\|_2,$$

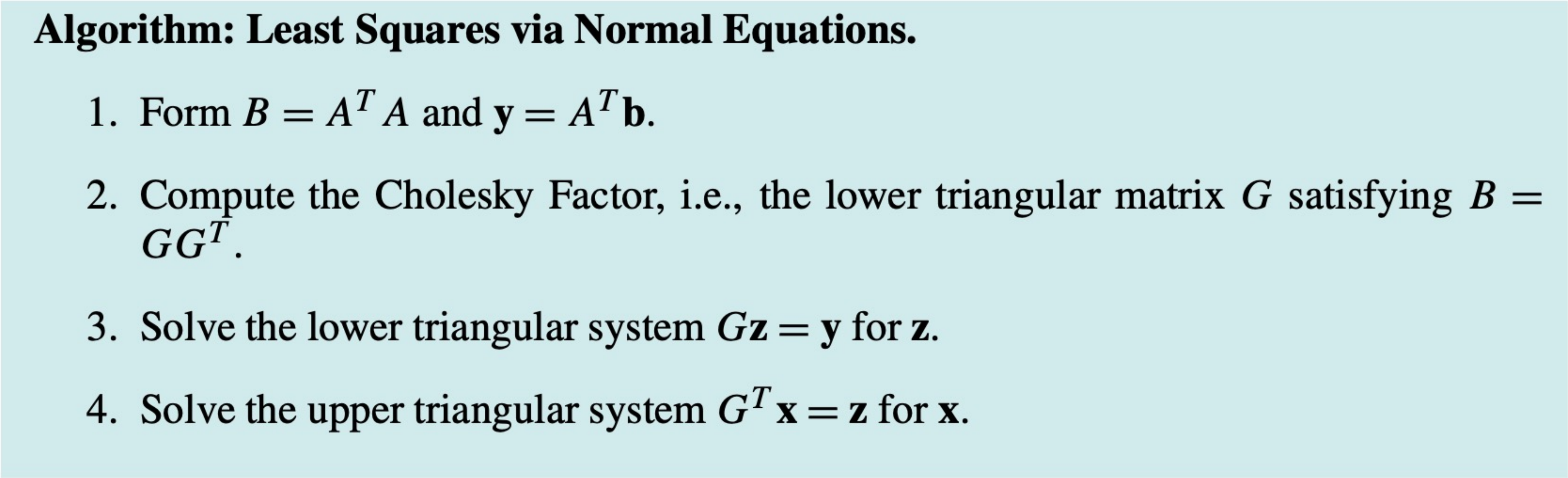
where A has full column rank, has a unique solution that satisfies the normal equations

$$\left(A^T A\right)\mathbf{x} = A^T \mathbf{b}.$$

In closed form, we use the **Pseudoinverse** of A: $A^\dagger = (A^T A)^{-1} A^T$

Inverting matrices is not preferred in practice. The system can be solved with LU Decomposition!

Least Squares Algorithm



Algorithm: Least Squares via Normal Equations.

1. Form $B = A^T A$ and $\mathbf{y} = A^T \mathbf{b}$.
2. Compute the Cholesky Factor, i.e., the lower triangular matrix G satisfying $B = GG^T$.
3. Solve the lower triangular system $G\mathbf{z} = \mathbf{y}$ for $\mathbf{z}$.
4. Solve the upper triangular system $G^T \mathbf{x} = \mathbf{z}$ for $\mathbf{x}$.

Linear Least Squares Example

$$\begin{matrix} 2 & -1 \\ 1 & 2 \\ 1 & 1 \end{matrix} * \begin{matrix} x \\ y \end{matrix} = \begin{matrix} 2 \\ 1 \\ 4 \end{matrix}$$

Let’s try to solve:

$$\left.\begin{matrix} 2 & -1 \\ 1 & 2 \\ 1 & 1 \end{matrix}\,\right|\,\begin{matrix} 2 \\ 1 \\ 4 \end{matrix} \quad \Rightarrow \quad \left.\begin{matrix} 1 & -1 \\ 0 & 2.5 \\ 0 & 1.5 \end{matrix}\,\right|\,\begin{matrix} 2 \\ 0 \\ 3 \end{matrix}$$

R2: y = 0
R3: y = 2
Inconsistent system!

We have to use least squares method.

Step 1: Form $B = A^T A$ and $y = A^T b$

B: $$\begin{matrix} 2 & 1 & 1 \\ -1 & 2 & 1 \end{matrix} * \begin{matrix} 2 & -1 \\ 1 & 2 \\ 1 & 1 \end{matrix} = \begin{matrix} 6 & 1 \\ 1 & 6 \end{matrix}$$

Y: $$\begin{matrix} 2 & 1 & 1 \\ -1 & 2 & 1 \end{matrix} * \begin{matrix} 2 \\ 1 \\ 4 \end{matrix} = \begin{matrix} 9 \\ 4 \end{matrix}$$

Linear Least Squares Solution

Step 2: Complete the Cholesky Decomposition of B:

$$\begin{matrix} 6 & 1 \\ 1 & 6 \end{matrix} \Rightarrow \quad L = \begin{matrix} l_{11} & l_{12} \\ l_{21} & l_{22} \end{matrix} = \begin{matrix} \sqrt{b_{11}} & 0 \\ \frac{b_{21}}{l_{11}} & \sqrt{b_{22} - l_{21}^2} \end{matrix} = \begin{matrix} \sqrt{6} & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{6 - \frac{1}{\sqrt{6}}^2} \end{matrix} = \begin{matrix} \sqrt{6} & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{6 - \frac{1}{6}} \end{matrix} = \begin{matrix} \sqrt{6} & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{\frac{35}{6}} \end{matrix}$$

$$\boxed{G = \begin{matrix} \sqrt{6} & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{\frac{35}{6}} \end{matrix}}$$

$$\boxed{G^T = \begin{matrix} \sqrt{6} & \frac{1}{\sqrt{6}} \\ 0 & \sqrt{\frac{35}{6}} \end{matrix}}$$

Linear Least Squares Solution

<u>Step 3:</u> Solve $Gz = \boldsymbol{y}$ for **z** (Use Forward Substitution)

$$G = \begin{matrix} \sqrt{6} & 0 \\ \frac{1}{\sqrt{6}} & \sqrt{\frac{35}{6}} \end{matrix} \quad y = \begin{matrix} 9 \\ 4 \end{matrix} \quad z = \begin{matrix} z_1 \\ z_2 \end{matrix}$$

$$\boxed{z = \begin{matrix} \frac{9}{\sqrt{6}} \\ \sqrt{\frac{15}{14}} \end{matrix}}$$

$$z_1 = \frac{9}{\sqrt{6}}$$

$$z_2 = \frac{4 - \frac{1}{\sqrt{6}}\frac{9}{\sqrt{6}}}{\sqrt{\frac{35}{6}}} = \frac{\frac{24}{6} - \frac{9}{6}}{\sqrt{\frac{35}{6}}} = \frac{\frac{15}{6}}{\sqrt{\frac{35}{6}}} = \frac{15}{6}\sqrt{\frac{6}{35}} = \sqrt{\frac{15}{14}}$$

Linear Least Squares Solution

Step 4: Solve $G^T\boldsymbol{x} = \boldsymbol{z}$ for **x** (Use Backward Substitution)

$$G^T = \begin{matrix} \sqrt{6} & \frac{1}{\sqrt{6}} \\ 0 & \sqrt{\frac{35}{6}} \end{matrix} \quad z = \begin{matrix} \frac{9}{\sqrt{6}} \\ \sqrt{\frac{15}{14}} \end{matrix} \quad x = \begin{matrix} x_1 \\ x_2 \end{matrix}$$

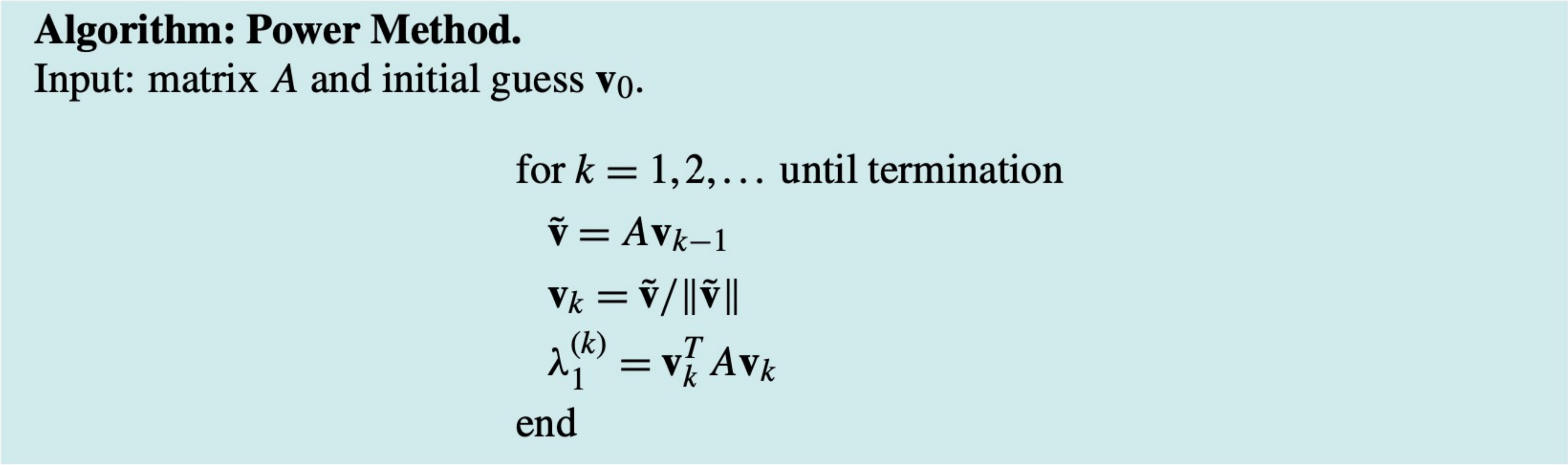
$$\boxed{x = \begin{matrix} \frac{10}{7} \\ \frac{3}{7} \end{matrix}}$$

$$x_2 = \frac{\sqrt{\frac{15}{14}}}{\sqrt{\frac{35}{6}}} = \frac{3}{7} \qquad x_1 = \frac{\frac{9}{\sqrt{6}} - \frac{1}{\sqrt{6}}\frac{3}{7}}{\sqrt{6}} = \frac{10}{7}$$

CHP 8

Power Method

Use the Power Method to find the dominant eigenvector and the corresponding eigenvalue.

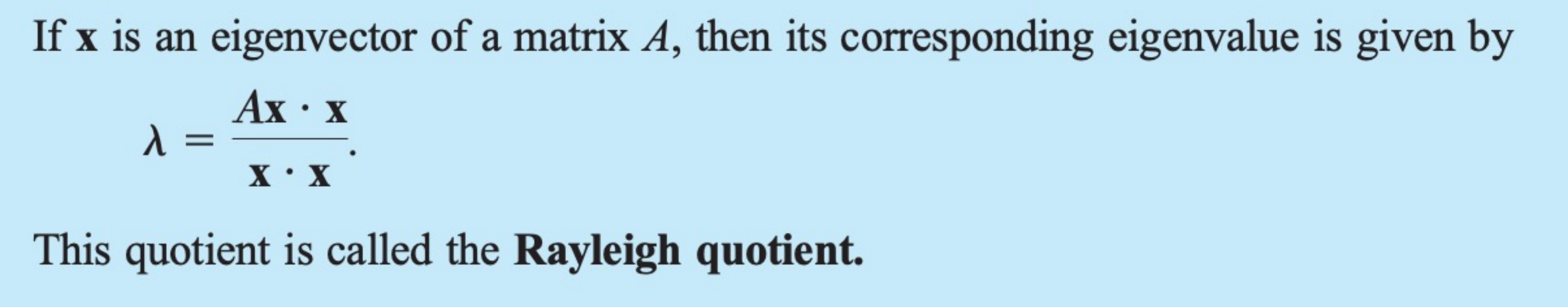


Algorithm: Power Method.
Input: matrix A and initial guess $\mathbf{v}_0$.

```
for $k = 1, 2, \ldots$ until termination
    $\tilde{\mathbf{v}} = A\mathbf{v}_{k-1}$
    $\mathbf{v}_k = \tilde{\mathbf{v}} / \|\tilde{\mathbf{v}}\|$
    $\lambda_1^{(k)} = \mathbf{v}_k^T A \mathbf{v}_k$
end
```

Rayleigh Quotient

To determine the eigenvalue for an eigenvector, use the Rayleigh quotient:



If $\mathbf{x}$ is an eigenvector of a matrix A, then its corresponding eigenvalue is given by

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}}.$$

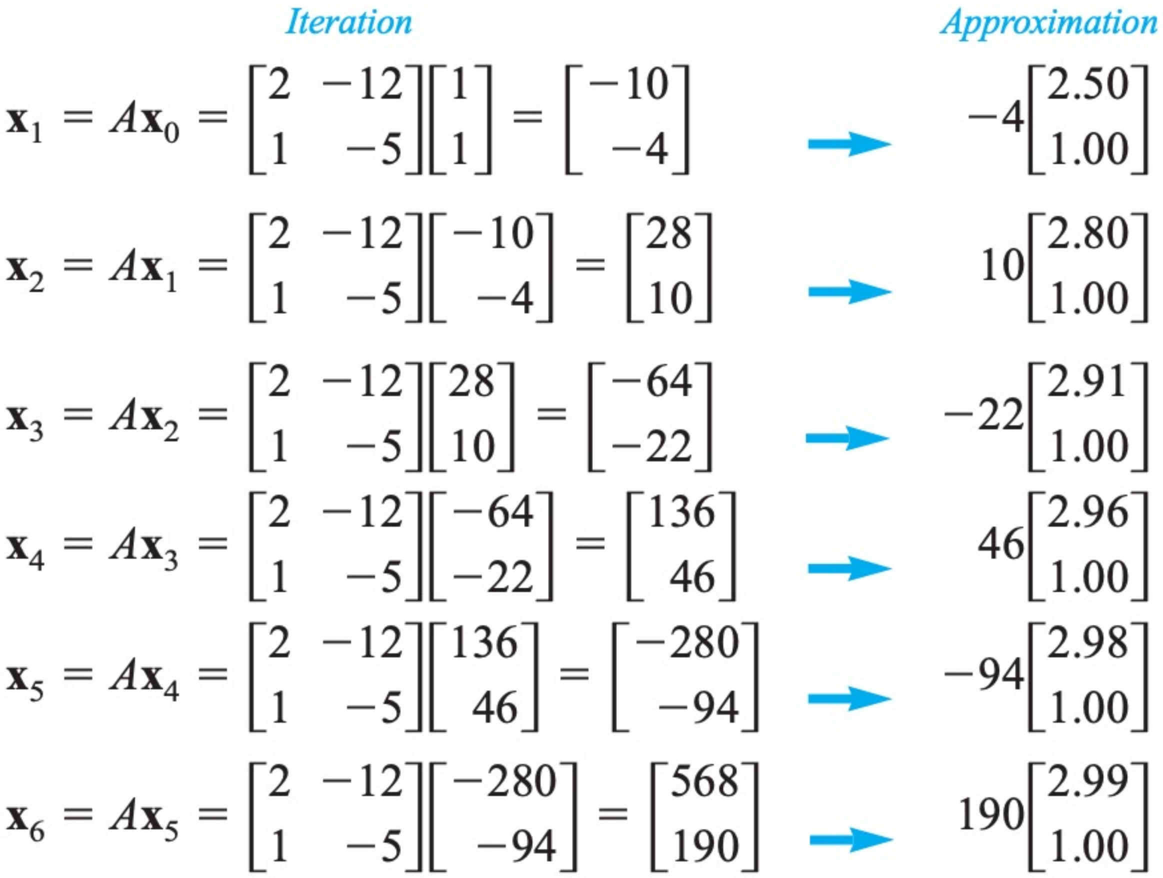
This quotient is called the **Rayleigh quotient.**

Power Method Example

- Complete 6 iterations of the Power Method to approximate the dominant eigenvector of the matrix given below. Use initial approximation x_0.

- $\mathbf{A} = \begin{vmatrix} 2 & -12 \\ 1 & -5 \end{vmatrix}$ $\quad\quad\quad \mathbf{x} = \begin{vmatrix} 1 \\ 1 \end{vmatrix}$

Power Method Solution



Iteration — *Approximation*

$$\mathbf{x}_1 = A\mathbf{x}_0 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -10 \\ -4 \end{bmatrix} \rightarrow -4\begin{bmatrix} 2.50 \\ 1.00 \end{bmatrix}$$

$$\mathbf{x}_2 = A\mathbf{x}_1 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -10 \\ -4 \end{bmatrix} = \begin{bmatrix} 28 \\ 10 \end{bmatrix} \rightarrow 10\begin{bmatrix} 2.80 \\ 1.00 \end{bmatrix}$$

$$\mathbf{x}_3 = A\mathbf{x}_2 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 28 \\ 10 \end{bmatrix} = \begin{bmatrix} -64 \\ -22 \end{bmatrix} \rightarrow -22\begin{bmatrix} 2.91 \\ 1.00 \end{bmatrix}$$

$$\mathbf{x}_4 = A\mathbf{x}_3 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -64 \\ -22 \end{bmatrix} = \begin{bmatrix} 136 \\ 46 \end{bmatrix} \rightarrow 46\begin{bmatrix} 2.96 \\ 1.00 \end{bmatrix}$$

$$\mathbf{x}_5 = A\mathbf{x}_4 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} 136 \\ 46 \end{bmatrix} = \begin{bmatrix} -280 \\ -94 \end{bmatrix} \rightarrow -94\begin{bmatrix} 2.98 \\ 1.00 \end{bmatrix}$$

$$\mathbf{x}_6 = A\mathbf{x}_5 = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix}\begin{bmatrix} -280 \\ -94 \end{bmatrix} = \begin{bmatrix} 568 \\ 190 \end{bmatrix} \rightarrow 190\begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}$$

(Approaching [3, 1])

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}}.$$

Power Method Solution

$$\mathbf{x}_6 = \begin{bmatrix} 568 \\ 190 \end{bmatrix} \approx 190 \begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix}$$

$$A\mathbf{x} = \begin{bmatrix} 2 & -12 \\ 1 & -5 \end{bmatrix} \begin{bmatrix} 2.99 \\ 1.00 \end{bmatrix} = \begin{bmatrix} -6.02 \\ -2.01 \end{bmatrix}$$

$$A\mathbf{x} \cdot \mathbf{x} = (-6.02)(2.99) + (-2.01)(1) \approx -20.0$$

$$\mathbf{x} \cdot \mathbf{x} = (2.99)(2.99) + (1)(1) \approx 9.94$$

$$\lambda = \frac{A\mathbf{x} \cdot \mathbf{x}}{\mathbf{x} \cdot \mathbf{x}} \approx \frac{-20.0}{9.94} \approx -2.01$$

Which is a good estimate of the dominant eigenvalue ($\lambda_1 = -2$)

Singular Value Decomposition

- Sometimes, a solution to the problem Ax = b is impossible to find because it is an ill conditioned number (i.e. condition number is very large).
- Thus, it is favorable to attempt to regularize it.
- This means, replace the given problem intelligently by a nearby problem which is better conditioned.
- Using SVD this can be done by setting the singular values below a cutoff tolerance to 0, and minimizing the l2-norm of the solution to the resulting underdetermined problem.

https://www.youtube.com/watch?v=mBcLRGuAFUk

Singular Value Decomposition

- A = PDP^{-1}
- P = Eigenvectors
- D = Diagonal Matrix of Eigenvalues
- Great, but not every matrix is diagonalizable.
- SVD: $A = U\Sigma V^T$
  - U: Orthogonal Left Singular Vectors
  - Σ: Diagonal Matrix of Singular Values
  - V^T:Orthogonal Right Singular Vectors
  - Advantage: you can do it to any matrix

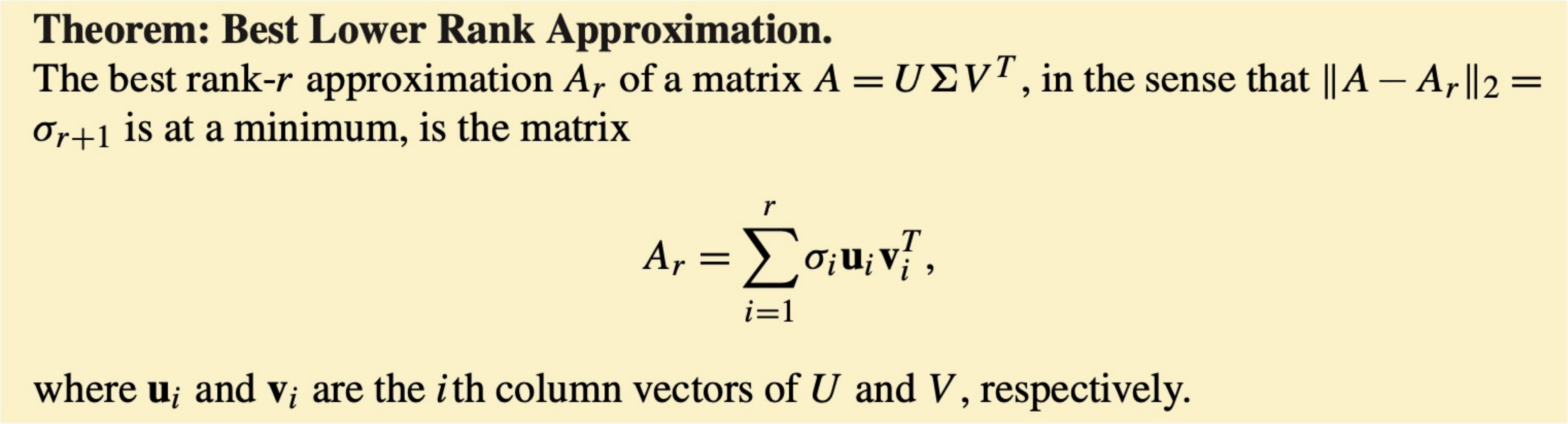
Singular Value Decomposition

- Suppose $A = U\Sigma V^T$
- $AA^T = (U\Sigma V^T)(U\Sigma V^T)^T$
- $AA^T = U\Sigma V^T V \Sigma^T U^T$
- $AA^T = U\Sigma^2 U^T$

- Σ^2= Eigenvalues of AA^T
- U & U^T= Eigenvectors of AA^T

- The same can be done for $A^T A$
  - Σ^2= Eigenvalues of$A^T A$
  - V & V^T= Eigenvectors of $A^T A$

Best Lower Rank Approximation



Theorem: Best Lower Rank Approximation.

The best rank-r approximation A_r of a matrix $A = U\Sigma V^T$, in the sense that $\|A - A_r\|_2 = \sigma_{r+1}$ is at a minimum, is the matrix

$$A_r = \sum_{i=1}^{r} \sigma_i \mathbf{u}_i \mathbf{v}_i^T,$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the ith column vectors of U and V, respectively.

SVD Example

- Find the SVD of $\begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix}$

SVD Solution

Step 1: Calculate AA^T and find Eigenvalues and Eigenvectors

$$\begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} \begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} = \begin{vmatrix} 8 & 4 \\ 4 & 2 \end{vmatrix}$$

Eigenvalues:

$$\lambda_1 = 10$$
$$\lambda_2 = 0$$

Eigenvectors:

$$\begin{array}{cc} v_1 & v_2 \\ \begin{vmatrix} 2 \\ 1 \end{vmatrix} & \begin{vmatrix} -1 \\ 2 \end{vmatrix} \end{array}$$

Step 2: Calculate $A^T A$and find Eigenvalues and Eigenvectors

$$\begin{vmatrix} 2 & 1 \\ 2 & 1 \end{vmatrix} \begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} =(5) \begin{vmatrix} 1 & 1 \\ 1 & 1 \end{vmatrix}$$

Eigenvalues:

$$\lambda_1 = 10$$
$$\lambda_2 = 0$$

Eigenvectors:

$$\begin{array}{cc} v_1 & v_2 \\ \begin{vmatrix} 1 \\ 1 \end{vmatrix} & \begin{vmatrix} 1 \\ -1 \end{vmatrix} \end{array}$$

SVD Solution

- $U = \frac{1}{\sqrt{5}} \begin{vmatrix} 2 & -1 \\ 1 & 2 \end{vmatrix}$
- $\Sigma \quad \begin{vmatrix} \sqrt{10} & 0 \\ 0 & 0 \end{vmatrix}$
- $V^T = \frac{1}{\sqrt{2}} \begin{vmatrix} 1 & 1 \\ 1 & -1 \end{vmatrix}$

https://www.youtube.com/watch?v=mBcLRGuAFUk&t=658s