University Course

Math 501
Advanced Numerical Analysis and Computing

California state university, Fullerton

Spring, 2007

My Class Notes

Nasser M. Abbasi

Summer 2021

Contents

Chapter 1

Introduction

Local contents

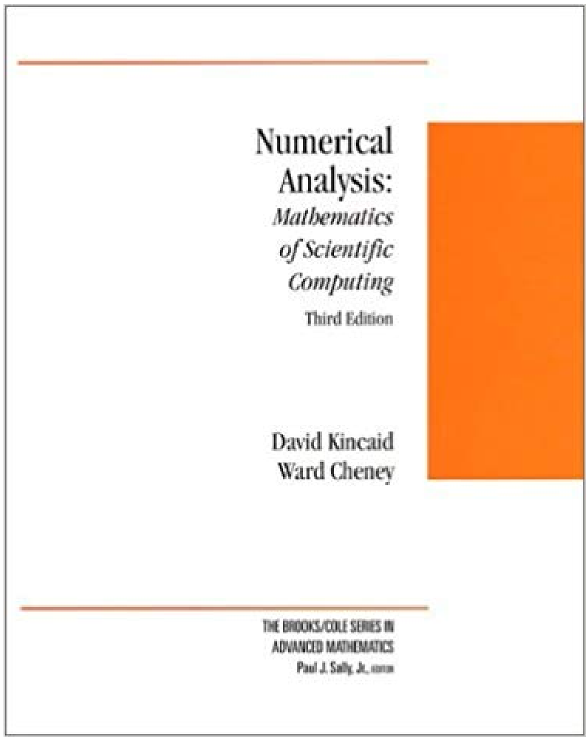
I took this course during summer 2007, at California state univ. Fullerton. This was a required course for my MSc. In Applied Mathematics. Instructor is

Dr Charles. H. Lee (Hung V. Ly),
Associate Professor
B.Sc., M.Sc., Ph.D., University of California, Irvine.

Department of Mathematics
California State University
Fullerton, CA 92834

Professor Lee's web site is `http://newton.fullerton.edu/hunglee/`

1.1 Text book



Numerical
Analysis:
Mathematics
of Scientific
Computing
Third Edition

David Kincaid
Ward Cheney

THE BROOKS/COLE SERIES IN
ADVANCED MATHEMATICS
Paul J. Sally, Jr., EDITOR

1.1.1 My review of the book

The book contains lots of technical information. Almost anything related to numerical analysis is contained in this book.

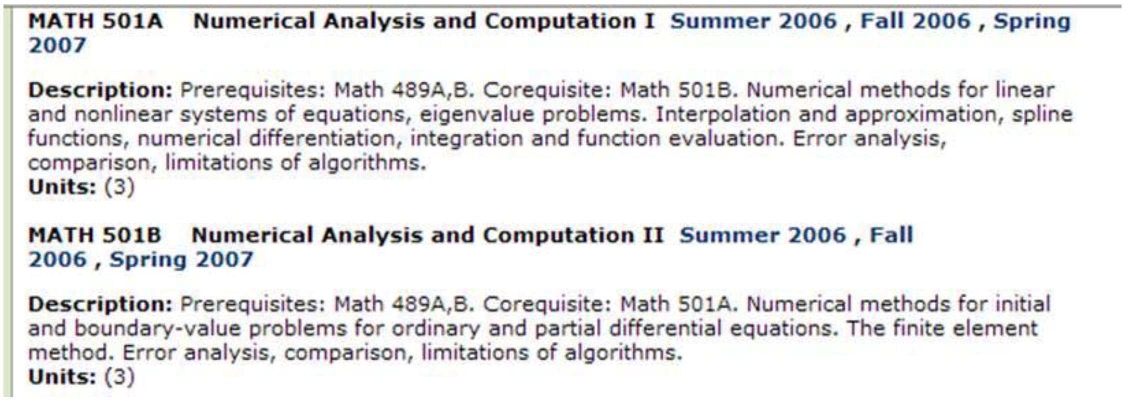
However, this book has a very dry style. It was hard for me to learn from it. It contains small number of worked out examples. The style of presentation is too formal and hence this text is best used as a reference more than something a student can learn from.

I found many other books which was from me easier to read and learn from.

Again, this is a good book, but as a reference.

This was a really hard course for me. Too many theorems and proofs to learn. Our instructor Dr Lee was very good and knew the material very well, but for me this was still a hard because it contained more real math than I am used to in the engineering courses I took before (but again, this is an applied math course, so what did I expect?)

1.2 course description



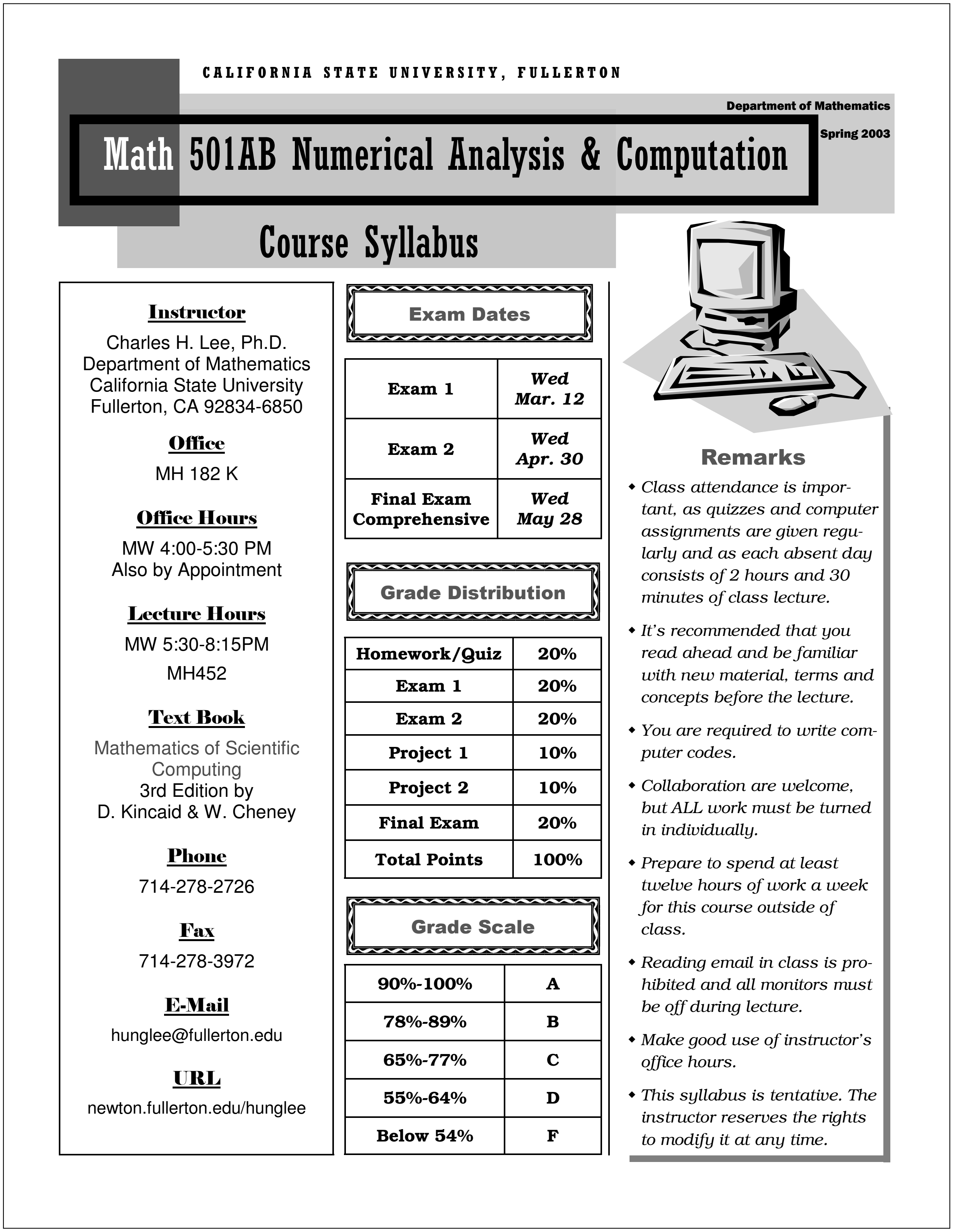
MATH 501A Numerical Analysis and Computation I Summer 2006 , Fall 2006 , Spring 2007

Description: Prerequisites: Math 489A,B. Corequisite: Math 501B. Numerical methods for linear and nonlinear systems of equations, eigenvalue problems. Interpolation and approximation, spline functions, numerical differentiation, integration and function evaluation. Error analysis, comparison, limitations of algorithms.
Units: (3)

MATH 501B Numerical Analysis and Computation II Summer 2006 , Fall 2006 , Spring 2007

Description: Prerequisites: Math 489A,B. Corequisite: Math 501A. Numerical methods for initial and boundary-value problems for ordinary and partial differential equations. The finite element method. Error analysis, comparison, limitations of algorithms.
Units: (3)

1.3 Syllabus



CALIFORNIA STATE UNIVERSITY, FULLERTON

Department of Mathematics

Spring 2003

Math 501AB Numerical Analysis & Computation

Course Syllabus

Instructor

Charles H. Lee, Ph.D.
Department of Mathematics
California State University
Fullerton, CA 92834-6850

Office

MH 182 K

Office Hours

MW 4:00-5:30 PM
Also by Appointment

Lecture Hours

MW 5:30-8:15PM
MH452

Text Book

Mathematics of Scientific Computing
3rd Edition by
D. Kincaid & W. Cheney

Phone

714-278-2726

Fax

714-278-3972

E-Mail

hunglee@fullerton.edu

URL

newton.fullerton.edu/hunglee

Exam Dates

Exam 1	*Wed Mar. 12*
Exam 2	*Wed Apr. 30*
Final Exam Comprehensive	*Wed May 28*

Grade Distribution

Homework/Quiz	20%
Exam 1	20%
Exam 2	20%
Project 1	10%
Project 2	10%
Final Exam	20%
Total Points	100%

Grade Scale

90%-100%	A
78%-89%	B
65%-77%	C
55%-64%	D
Below 54%	F

Remarks

- *Class attendance is important, as quizzes and computer assignments are given regularly and as each absent day consists of 2 hours and 30 minutes of class lecture.*
- *It's recommended that you read ahead and be familiar with new material, terms and concepts before the lecture.*
- *You are required to write computer codes.*
- *Collaboration are welcome, but ALL work must be turned in individually.*
- *Prepare to spend at least twelve hours of work a week for this course outside of class.*
- *Reading email in class is prohibited and all monitors must be off during lecture.*
- *Make good use of instructor's office hours.*
- *This syllabus is tentative. The instructor reserves the rights to modify it at any time.*

Chapter 2

my typed lecture notes

Local contents

2.1 Introduction

Each student had to typeset some lectures, here are mine

2.2 Lecture monday April 17 2007

2.2.1 Section 5.4

2.2.1.1 Theorem 1: Singular value decomposition

Given any arbitrary matrix $A_{m\times n}$ it can be factored into 3 matrices as follows $A_{m\times n} = P_{m\times m}D_{m\times n}Q_{n\times n}$ where P is a unitary matrix ($P^H = P^{-1}$ or $P^HP = I$), and Q is also unitary matrix.

These are the steps to do SVD

1. Find the rank of A, say r
2. Let $B = A^H_{n\times m}A_{m\times n}$, hence $B_{n\times n}$ is a square matrix, and it is semi positive definite, hence its eigenvalues will all be ≥ 0. Find the eigenvalues of B, call these σ_i^2. There will be n such eigenvalues since B is of order $n\times n$. But only r of these will be positive, and $n-r$ will be zero. Arrange these eigenvalues such that the first r non-zero eigenvalues come first, followed by the zero eigenvalues: $\overbrace{\sigma_1^2,\sigma_2^2,\cdots,\sigma_r^2,0,0,\cdots,0}^{n\ eigenvalues}$
3. Initialize matrix $D_{m\times n}$ to be all zeros. Take the the first r eigenvalues from above (non-zero ones), and take the square root of each, hence we get $\overbrace{\sigma_1,\sigma_2,\cdots,\sigma_r}^{r\text{ singular values}}$,and write these down the diagonal of D starting at $D(1,1)$, i.e. $D(1,1)=\sigma_1, D(2,2)=\sigma_2,\cdots,D(r,r)=\sigma_r$. Notice that the matrix D need not square matrix. Hence we can obtain an arrangement such as the following for $r=2$ $\begin{pmatrix}\sigma_1 & 0 & 0 & 0\\ 0 & \sigma_2 & 0 & 0\\ 0 & 0 & 0 & 0\end{pmatrix}$, where the matrix A was 3×4 for example.
4. Now find each eigenvalue of A^HA. For each eigenvalue, find the corresponding eigenvector. Call these eigenvectors $\vec{u}_1,\vec{u}_2,\cdots,\vec{u}_n$.
5. Normalize the eigenvectors found in the above step. Now $\vec{u}_1,\vec{u}_2,\cdots,\vec{u}_n$ eigenvector will be an orthonormal set of vectors. Take the Hermitian of each eigenvector $\vec{u}_1^H,\vec{u}_2^H,\cdots,\vec{u}_n^H$ and make one of these vectors (now in row format instead of column format) go into a row in the matrix Q.i.e the first row of Q will be $\vec{u}_1^H$, the second row of Q will be $\vec{u}_2^H$, etc... $\left(A^HA\right)_{n\times n}\xrightarrow{\text{find eigenvectors and normalize}}\left\{\vec{u}_1,\vec{u}_2,\cdots,\vec{u}_r,\quad\overbrace{\vec{u}_{r+1},\cdots,\vec{u}_n}^{\text{ortho basis (n-r) for N(A)}}\right\}\rightarrow$

$$Q_{n\times n} = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_r^T \\ \vec{u}_{r+1}^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix}$$

6. Now we need to find a set of m orthonormal vectors, these will be the columns of the matrix P. There are 2 ways to do this. First the textbook way, and then another way which I think is simpler.

   (a) The textbook method is as follows: find a set of r orthonormal eigenvector $\vec{v}_i = \frac{1}{\sigma_i} A\vec{u}_i$, for $i = 1 \cdots r$. Notice that here we only use the first r vectors found in step 5. Take each one of these $\vec{v}_i$ vectors and make them into columns of P. But since we need m columns in P not just r, we need to come up with $m - r$ more basis vectors such that all the m vectors form an orthonormal set of basis vectors for the row space of A, ie. $\mathbb{C}^m$. If doing this by hand, it is easy to find the this $m - r$ by inspection. In a program, we could use the same process we used with Gram-Schmidt, where we learned how find a new vector which is orthonormal to a an existing set of other vectors.

   (b) An easier approach is: Find AA^H (do not confused with how we did this in step 4 where we did A^HA). This will be an $m \times m$ matrix. Now find each eigenvalue. For each eigenvalue, find the corresponding eigenvector. Now normalize these basis vectors. These will now be an orthonormal eigenvectors $\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m$. Now as in the above step, these vectors will becomes the columns of the matrix P. The difference in this approach is that we do not need to use Gram-Schmidt to find the $m-r$ eigenvectors since we will obtain the m eigenvectors right away.

   $$\left(AA^H\right)_{m\times m} \xrightarrow{\text{find eigenvectors and normalize}} \left\{ \overbrace{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_r}^{\text{r ortho basis for range A}}, \vec{v}_{r+1}, \cdots, \vec{v}_m \right\} \rightarrow P_{m\times m} =$$

   $$\begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_m \end{bmatrix}$$

7. This completes the factorization, now we have $A = PDQ$

In Matlab, to find SVD, use the command $[P, D, Q] = svd(A)$.

The following diagram help illustrate the SVD process described above.

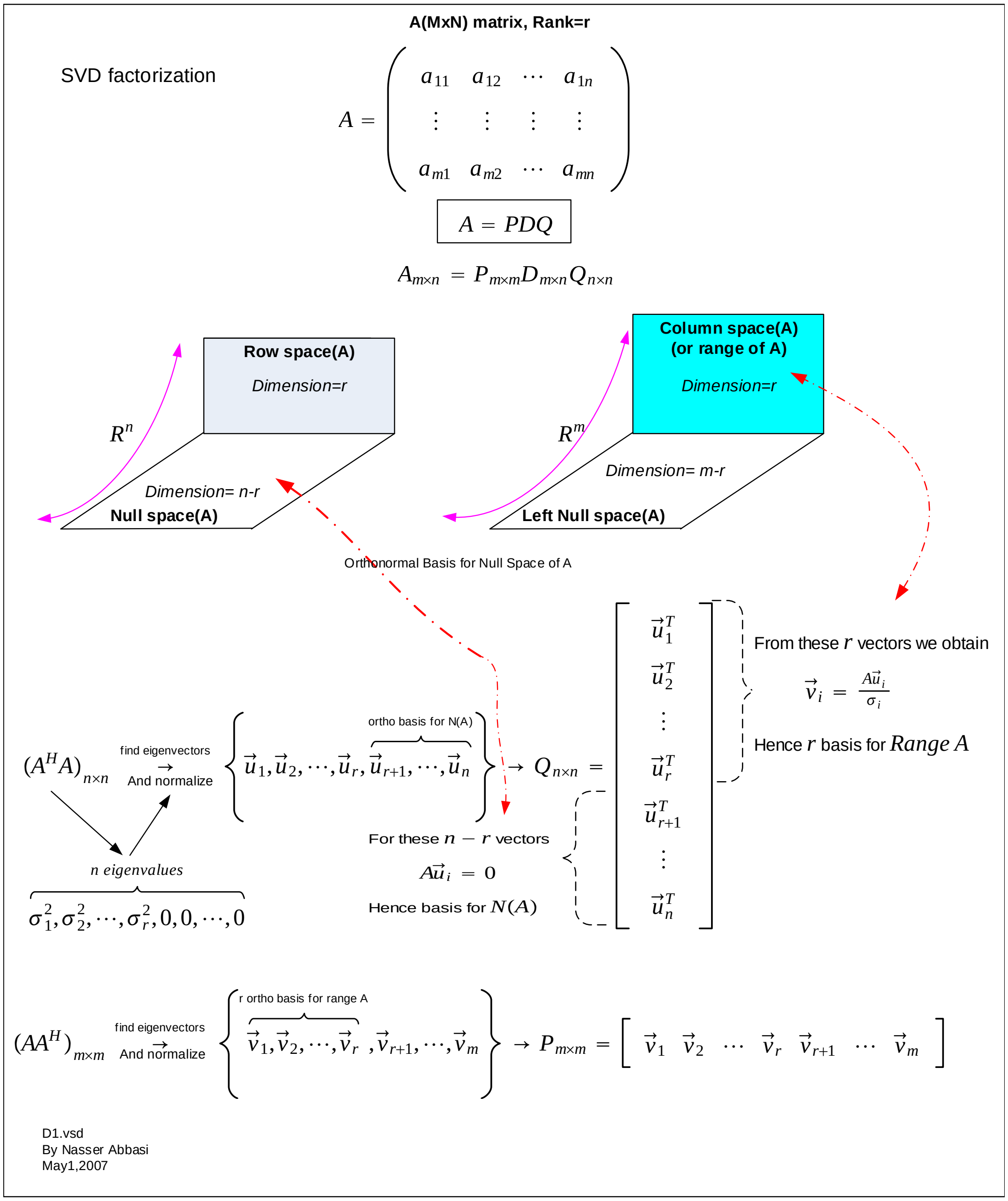

A(MxN) matrix, Rank=r
SVD factorization
$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$
$\boxed{A = PDQ}$
$A_{m\times n} = P_{m\times m} D_{m\times n} Q_{n\times n}$
Row space(A)
Dimension=r
R^n
Dimension= n-r
Null space(A)
Column space(A)
(or range of A)
Dimension=r
R^m
Dimension= m-r
Left Null space(A)
Orthonormal Basis for Null Space of A
$(A^H A)_{n\times n}$
find eigenvectors
And normalize
ortho basis for N(A)
$\left\{ \vec{u}_1, \vec{u}_2, \cdots, \vec{u}_r, \overbrace{\vec{u}_{r+1}, \cdots, \vec{u}_n} \right\} \to Q_{n\times n} =$
$\begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \\ \vdots \\ \vec{u}_r^T \\ \vec{u}_{r+1}^T \\ \vdots \\ \vec{u}_n^T \end{bmatrix}$
From these r vectors we obtain
$\vec{v}_i = \frac{A\vec{u}_i}{\sigma_i}$
Hence r basis for $Range\ A$
For these $n - r$ vectors
$A\vec{u}_i = 0$
Hence basis for $N(A)$
n eigenvalues
$\overbrace{\sigma_1^2, \sigma_2^2, \cdots, \sigma_r^2, 0, 0, \cdots, 0}$
$(AA^H)_{m\times m}$
find eigenvectors
And normalize
r ortho basis for range A
$\left\{ \overbrace{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_r}, \vec{v}_{r+1}, \cdots, \vec{v}_m \right\} \to P_{m\times m} = \begin{bmatrix} \vec{v}_1 & \vec{v}_2 & \cdots & \vec{v}_r & \vec{v}_{r+1} & \cdots & \vec{v}_m \end{bmatrix}$
D1.vsd
By Nasser Abbasi
May1,2007

Figure 2.1: SVD

Remarks:

1. $\sigma_1, \sigma_2, \cdots, \sigma_n$ are called the singular values of A.
2. the SVD factorization is not unique as $\sigma_1, \sigma_2, \cdots, \sigma_r$ can be ordered differently. Also the choice of $m - r$ orthonormal basis for the P matrix is arbitrary.

2.2.1.2 Pseudo Inverse

Given an arbitrary matrix A to find its pseudo inverse, called A^{+} we start by first find the SVD of A, then write

$$A^{+} = Q^{H} D^{+} P^{H} \tag{1}$$

Where, since D is diagonal $\begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \ddots \end{pmatrix}$, then $D^{+} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & 0 \\ 0 & \frac{1}{\sigma_2} & 0 \\ 0 & 0 & \ddots \end{pmatrix}$, and so finding the pseudo inverse of a matrix is straightforward if we have the SVD factorization of the matrix. To show equation (1): Since

$$A = PDQ$$

Then pre multiply both sides by P^{H} we obtain

$$P^{H} A = P^{H} P D Q$$

But $P^{H} = P^{-1}$ since P is unitary matrix, then the above becomes

$$P^{H} A = DQ$$

Now post-multiply both sides by Q^{H}

$$P^{H} A Q^{H} = D Q Q^{H}$$

But $Q^{H} = Q^{-1}$ since Q is unitary matrix, then the above becomes

$$P^{H} A Q^{H} = D$$

Hence

$$D^{+} = \left(P^{H} A Q^{H}\right)^{+}$$

To find definition for A^{+}, start from $Ax = b$

$$\begin{aligned} Ax &= b \\ A^{H} A x &= A^{H} b \\ x &= \left(A^{H} A\right)^{-1} A^{H} b \end{aligned}$$

Hence

$$\begin{aligned}
A^{+} &= \left(A^{H}A\right)^{-1} A^{H} \\
&= \left((PDQ)^{H}(PDQ)\right)^{-1}(PDQ)^{H} \\
&= \left(\left(Q^{H}D^{H}P^{H}\right)(PDQ)\right)^{-1}\left(Q^{H}D^{H}P^{H}\right) \\
&= \left((PDQ)^{-1}\left(Q^{H}D^{H}P^{H}\right)^{-1}\right)\left(Q^{H}D^{H}P^{H}\right) \\
&= \left(\left(Q^{-1}D^{-1}P^{-1}\right)\left(P^{-H}D^{-H}Q^{-H}\right)\right)\left(Q^{H}D^{H}P^{H}\right) \\
&= \left(Q^{-1}D^{-1}P^{-1}P^{-H}D^{-H}Q^{-H}\right)\left(Q^{H}D^{H}P^{H}\right) \\
&= \left(Q^{-1}D^{-1}D^{-H}Q^{-H}\right)\left(Q^{H}D^{H}P^{H}\right) \\
&= Q^{-1}D^{-1}D^{-H}Q^{-H}Q^{H}D^{H}P^{H} \\
&= Q^{-1}D^{-1}D^{-H}D^{H}P^{H} \\
&= Q^{-1}D^{-1}P^{H} \\
&= Q^{H}D^{-1}P^{H}
\end{aligned}$$

But $D^{-1} = D^{+}$ hence the above becomes

$$A^{+} = Q^{H}D^{+}P^{H}$$

which is equation (1).

Note: $A^{+}A = I^{+}$, where I^{+} has the form $\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$. in other words, it does not have 1 on all the diagonal elements as the normal I would.

2.2.1.3 Theorem 2: Minimal solution to $A\vec{x} = \vec{b}$ and the pseudo inverse A^{+}

if $\hat{x} = A^{+}\vec{b}$ then $\hat{x}$ is the minimal solution of $A\vec{x} = \vec{b}$

proof:

let $c = P^{H}b$, $y = Qx$

$$
\begin{aligned}
\rho &= \inf_x \|Ax-b\|_2 \\
&= \inf_x \|PDQx-b\|_2 \\
&= \inf_x \left\|P^H PDQx - P^H b\right\|_2 \\
&= \inf_x \left\|DQx - P^H b\right\|_2 \\
&= \inf_x \left\|Dy - c\right\|_2
\end{aligned}
$$

Since $D = \begin{pmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \sigma_r & \\ & & & 0 \end{pmatrix}$ then

$$\left\|Dy-c\right\|_2^2 = \sum_{i=1}^{r} \left(\sigma_i y_i - c_i\right)^2 + \sum_{i=r+1}^{n} c_i^2$$

This is minium when $y_i = \frac{\sigma_i}{c_i}$,where $i = 1 \cdots r$, then will make the first term in the RHS above go to zero. Hence

$$\rho = \sqrt{\sum_{i=r+1}^{n} c_i^2}$$

The vector $\vec{y}$ which gives this minium solution has $y_{r+1}, y_{r+2}, \cdots, y_n = 0$, hence

$$y = D^+ c$$

But since $y = Qx$, hence $D^+ c = Qx$ or $x = Q^H D^+ c$, but $c = P^H b$, then

$$
\begin{aligned}
x &= Q^H D^+ P^H b \\
&= A^+ b
\end{aligned}
$$

Hence

$$A^+ = Q^H D^+ P^H$$

2.2.1.4 Theorem 3: Penrose properties for any matrix A

For any $m \times n$ matrix A there is at least one matrix X with the following 4 properties

1. $AXA = A$
2. $XAX = X$
3. $(AX)^H = AX$
4. $(XA)^H = XA$

We need to show that the matrix X is unique (with respect to A). The proof is by contradiction. We assume that A has associated with it 2 matrices with the above properties, X and Y and we will show that $X = Y$, hence X is unique for A.

The proof starts by saying $X = XAX$ and through number of substitution steps using the above 4 we obtain Y as follows

$$\begin{aligned}
X &= X \overbrace{A}^{AYA} X \\
&= XAYAX \\
&= X \overbrace{A}^{AYA} Y \overbrace{A}^{AYA} X \\
&= XAYAYAYAX \\
&= (XA)^H (YA)^H Y (AY)^H (AX)^H \quad \text{using property 3,4} \\
&= A^H X^H A^H Y^H Y Y^H A^H X^H A^H \\
&= (AXA)^H Y^H Y Y^H (AXA)^H \\
&= \overbrace{(AXA)}^{H} Y^H Y Y^H \overbrace{(AXA)}^{H} \\
&= (A)^H Y^H Y Y^H (A)^H \qquad \text{property 1} \\
&= (YA)^H Y (AY)^H \\
&= YAYAY \quad \text{property 4 and 3} \\
&= YAY \quad \text{property 2} \\
&= Y \quad \text{property 2}
\end{aligned}$$

Hence $X = Y$, so X is unique.

2.2.1.5 Theorem 4: A^+ satisfies Penrose properties hence unique

Theorem: The pseudo inverse of a matrix has four Penrose properties. hence each matrix has a unique pseudo inverse A^+

proof: Let A be any matrix, its SVD is $A = PDQ$, and we showed before that $A^+ = Q^H D^+ P^H$

note: $DD^+D = D$

Let us now show that each property of Penrose is satisfied. In this case X is our A^+. Hence for property 1 we need to show that $AA^+A = A$.

$$\begin{aligned}
AA^+A &= PDQ\overbrace{Q^H D^+ P^H}^{A^+}PDQ \\
&= PDD^+P^HPDQ \\
&= PDD^+DQ \\
&= PDI^+Q \\
&= PDQ \\
&= A
\end{aligned}$$

To show property 2, we need to show that $A^+AA^+ = A^+$

$$\begin{aligned}
A^+AA^+ &= \left(Q^H D^+ P^H\right) A \left(Q^H D^+ P^H\right) \\
&= \left(Q^H D^+ P^H\right) PDQ \left(Q^H D^+ P^H\right) \\
&= \left(Q^H D^+\right) D \left(D^+ P^H\right) \\
&= Q^H D^+ D D^+ P^H \\
&= Q^H I^+ D^+ P^H \\
&= Q^H D^+ P^H \\
&= A^+
\end{aligned}$$

To show property 3, need to show $(AA^+)^H = AA^+$

$$\begin{aligned}
\left(AA^{+}\right)^{H} &= \left(\left(PDQ\right)\left(Q^{H}D^{+}P^{H}\right)\right)^{H}\\
&= \left(\left(Q^{H}D^{+}P^{H}\right)^{H}\left(PDQ\right)^{H}\right)\\
&= \left(P\left(D^{+}\right)^{H}Q\left(PDQ\right)^{H}\right)\\
&= \left(P\left(D^{+}\right)^{H}Q\left(Q^{H}D^{H}P^{H}\right)\right)\\
&= P\left(D^{+}\right)^{H}QQ^{H}D^{H}P^{H}\\
&= PD^{+}QQ^{H}D^{H}P^{H}\\
&= PD^{+}QQ^{H}D^{H}P^{H}\\
&= PDQQ^{H}D^{+}P^{H}\\
&= \left(PDQ\right)\left(Q^{H}D^{+}P^{H}\right)\\
&= AA+
\end{aligned}$$

For property (4) need to show $\left(A^{+}A\right)^{H} = A^{+}A$

$$\begin{aligned}
\left(A^{+}A\right)^{H} &= \left(\left(Q^{H}D^{+}P^{H}\right)\left(PDQ\right)\right)^{H}\\
&= \left(PDQ\right)^{H}\left(Q^{H}D^{+}P^{H}\right)^{H}\\
&= Q^{H}D^{H}P^{H}P\left(D^{+}\right)^{H}Q\\
&= Q^{H}DP^{H}P\left(D^{+}\right)Q\\
&= Q^{H}D\left(D^{+}\right)Q\\
&= Q^{H}D^{+}DQ\\
&= Q^{H}D^{+}P^{H}PDQ\\
&= A^{+}A
\end{aligned}$$

Hence Pseudocode is unique.

2.2.1.6 Theorem 5: On SVD properties

Let A have svd $A = PDQ$, then the following is true

1. rank of A is r
2. $\left\{\vec{u}_{r+1},\vec{u}_{r+2},\cdots,\vec{u}_{n}\right\}$ is an orthonormal base for null space of A
3. $\left\{\vec{v}_{1},\vec{v}_{2},\cdots,\vec{v}_{r}\right\}$ is an orthonormal base for range of of A
4. $\|A\|_{2} = \max_{1\leq i\leq n}|\sigma_{i}|$

proof:

1. Since $A = PDQ$ and since P, Q are invertible matrices (by construction, they are made up of eigenvectors from a set of distinct eigenvalues, hence they must be orthogonal eigenvectors). Therefor, the rank of A is decided by the rank of D. But D is all zeros except for the values σ_i each on a separate column and on separate row. Since there are r distinct values of those, hence D has r vectors that are linearly independent. Hence the rank of A is r.

2. Since $A\vec{u}_i = \vec{0}$ for each vector $i = (r+1)\cdots n$, and these $\vec{u}_i$ vectors form an orthogonal set of size $n-r$, and since the dimension of the null-space(A) is $n-r$ hence they span the whole null-space(A). Then they can be used as a basis for the null space of A (recall, the null-space of A is the space in which $A\vec{x} = \vec{0}$)

3. Since $\vec{v}_i = \frac{A\vec{u}_i}{\sigma_i} \neq \vec{0}$, for $i = 1\cdots r$, and these $\vec{v}_i$ vectors form an orthogonal set of basis, and since the dimension of the Range(A) is r hence they span the whole Range(A). Then they can be used as a basis for the Range(A).Another proof of this, is to use the construction of the $\vec{v}_i$ vector from AA^H. But we did not use this method, so I will leave this out for now.

4.
$$\begin{aligned}
\|A\|_2 &= \sup_{\|x\|_2=1}\left\{\left\|A\vec{x}\right\|_2\right\}\\
&= \sup_{\|x\|_2=1}\left\|PDQ\vec{x}\right\|_2\\
&= \sup_{\|y\|_2=1}\left\|D\vec{y}\right\|_2\\
&= \sup_{\|y\|_2=1}\sqrt{\sum_{i=1}^{r}\left(\sigma_i y_i\right)^2}\\
&= \max_{1\leq i\leq r}|\sigma_i|
\end{aligned}$$

2.2.1.7 Theorem 6: Reduced (or economical) SVD

Recall in the above standard SVD that when we build P matrix, we used the first r vectors $\vec{u}_i$ to generate $\vec{v}_i$ from, by doing $\vec{v}_i = \frac{A\vec{u}_i}{\sigma_i}$, then we used these $\vec{v}_i$ as the columns of P, but since P was of size $m \times m$ we had to come up with $m-r$ more orthogonal vectors to fill in the P matrix. In the economical SVD, we stop when we fill in r vectors in P. Hence P will be of size $m \times r$. Similarly for D we stop at the r singular value. Hence in this case D will be a diagonal matrix $r \times r$. And for the Q matrix, we also generate r orthonormal basis $\vec{u}_i$

Hence the economical SVD follows the same steps as the normal SVD, except we stop when we obtain r basis. The matrices will therefor be smaller, hence the name economical. The factorization will then be $A_{m\times n} = P_{m\times r}D_{r\times r}Q_{r\times n}$ (for the case when $m \geq n \geq r$)

I am not sure now in what application one would have to generate the full SVD vs. the economical SVD. But in Matlab, there is an optional to select either factorization. Type svd(A,'econ').

2.2.1.8 Theorem 7: On orthonormal Bases

Let L be a linear transformation from $\mathbb{C}^m$ to $\mathbb{C}^n$ then there are orthonormal bases $\vec{u}_1, \vec{u}_2, \cdots, \vec{u}_n$, for $\mathbb{C}^m$ and $v_1, v_2, \cdots, v_n$ for $\mathbb{C}^n$ s.t. $L\vec{u}_i = \begin{cases} \sigma_i \vec{v}_i & 1 \leq i \leq \min(m,n) \\ 0 & \min(m,n) < i \leq m \end{cases}$

proof: The proof as given in the textbook for this theorem is clear enough as outlined. Please see page 295.

2.2.2 Homework Solution for section 5.4

2.2.2.1 Problem section 5.4, 2(a)

question: Find the minimal solution for $x_1 x_2 = b$

answer:

First write the problem as

$$\begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = [b]$$

Minimal solution is $\vec{x} = A^+ b$, so we need to find A^+. Find $A = PDQ$, then $A^+ = Q^H D^+ P^H$

First find the set of $\vec{u}_i$ vectors to go to the Q matrix. I will use the economical SVD method.

$$\begin{aligned} A^H A &= \begin{bmatrix} 1 \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \end{aligned}$$

Hence $r = 1$

Hence $|A - \lambda I| = \begin{vmatrix} 1-\lambda & 1 \\ 1 & 1-\lambda \end{vmatrix} = 0 \rightarrow (1-\lambda)^2 - 1 = 0 \rightarrow 1 + \lambda^2 - 2\lambda - 1 = 0 \rightarrow \lambda(\lambda - 2) = 0$

Hence $\lambda_1 = 2, \lambda_2 = 0 \rightarrow \sigma_1 = \sqrt{2}, \sigma_2 = 0$

Find eigenvectors $\vec{u}_1, \vec{u}_2$.

For $\lambda_1 = 2 \rightarrow \begin{bmatrix} 1-2 & 1 \\ 1 & 1-2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\rightarrow \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow -y_1 + y_2 = 0 \Rightarrow \vec{u}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \overset{\text{normalize norm 2}}{\rightarrow} \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

For $\lambda_2 = 0 \rightarrow \begin{bmatrix} 1-0 & 1 \\ 1 & 1-0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

$\rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \Rightarrow y_1 + y_2 = 0 \Rightarrow \vec{u}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \overset{\text{normalize norm 2}}{\rightarrow} \begin{bmatrix} \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$

Hence $Q = \begin{bmatrix} \vec{u}_1^T \\ \vec{u}_2^T \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix} \frac{1}{\sqrt{2}}$

Not to find the P matrix. $AA^H = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = [2]$

The eigenvalue is $2 - \lambda = 0 \rightarrow \lambda = 2$. Hence the eigenvector is $2y_1 = 0 \rightarrow y_1$ =anything→ [1]

Hence the P matrix is [1]

The D matrix is $m \times n$, hence 1×2, then $D = \begin{bmatrix} \sigma_1 & 0 \end{bmatrix}$

Hence this completes the SVD. We have that

$$\begin{aligned}\begin{bmatrix}1 & 1\end{bmatrix} &= [1]\begin{bmatrix}\sigma_1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= [1]\begin{bmatrix}\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}\sqrt{2} & \sqrt{2}\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}1 & 1\end{bmatrix}\end{aligned}$$

So the SVD is verified. Not find

$$\begin{aligned}A^{+} &= Q^{H}D^{+}P^{H}\\ &= \begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}^{H}\frac{1}{\sqrt{2}}\begin{bmatrix}\sqrt{2} & 0\end{bmatrix}^{+}[1]^{H}\\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ 0\end{bmatrix}[1]\end{aligned}$$

Notice that D^{+} mean we also take the conjugate transpose of D and then we take the reciprocal of each entry. Hence if D is $m \times n$ then D^{+} is $n \times m$

$$\begin{aligned}A^{+} &= \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\end{aligned}$$

Hence

$$\begin{aligned}\hat{x} &= A^{+}b\\ &= \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix} b\end{aligned}$$

So the minimal solution is $x_1 = \frac{b}{2}, x_2 = \frac{b}{2}$

2.2.2.2 Problem 5.4, 10

Problem: Prove the following properties of A^{+}

a)$A^{++} = A$

b)$A^{+H} = A^{H+}$

answer:

a)$(A^{+})^{+} = \left((PDQ)^{+}\right)^{+} = \left(Q^{H}D^{+}P^{H}\right)^{+} = \left(P^{H}\right)^{H}(D^{+})^{+}\left(Q^{H}\right)^{H}$

But $\left(P^{H}\right)^{H} = P$ and $\left(Q^{H}\right)^{H} = Q$, and the reciprocal of a reciprocal gives us back the original value, hence $(D^{+})^{+} = D$

Hence we have $\left(P^{H}\right)^{H}(D^{+})^{+}\left(Q^{H}\right)^{H} = PDQ = A$

b)

$$\begin{aligned}(A^{+})^{H} &= \left(Q^{H}D^{+}P^{H}\right)^{H}\\ &= \left(P^{H}\right)^{H}(D^{+})^{H}\left(Q^{H}\right)^{H}\\ &= P(D^{+})^{H}Q \qquad (1)\end{aligned}$$

and

$$\begin{aligned}A^{H+} &= \left((PDQ)^{H}\right)^{+}\\ &= \left(Q^{H}D^{H}P^{H}\right)^{+}\\ &= \left(P^{H}\right)^{H}\left(D^{H}\right)^{+}\left(Q^{H}\right)^{H}\\ &= P\left(D^{H}\right)^{+}Q \qquad (2)\end{aligned}$$

Hence (1)=(2) if $D^{H+} = D^{+H}$. But this is the case. Since $D^{H+} = D$ and $D^{+H} = D$

2.2.2.3 Problem 5.4, 23

Problem: Let A be a square matrix with SVD PDQ, prove that the characteristic polynomial of $A = \pm\det\left(D - \lambda P^H Q^H\right) = 0$

answer:

Let the characteristic polynomial of A be called Δ, hence we write

$$\begin{aligned}
\Delta &= \det(A - \lambda I) \\
&= \det(PDQ - \lambda I) \\
&= \det\left(P\left(DQ - \lambda P^{-1}\right)\right) \\
&= \det\left(P\left(D - \lambda P^{-1}Q^{-1}\right)Q\right) \\
&= \det(P)\det\left(D - \lambda P^{-1}Q^{-1}\right)\det(Q)
\end{aligned}$$

But $Q^{-1} = Q^H$ and $P^{-1} = P^H$ hence the above becomes

$$\Delta = \det(P)\det\left(D - \lambda P^H Q^H\right)\det(Q)$$

Now $\Delta = 0$, and also we know that $\det(P) \neq 0$ and $\det(Q) \neq 0$, this is becuase P and Q are unitary matrices. Hence this means that

$$\det\left(D - \lambda P^H Q^H\right) = 0$$

2.2.2.4 Problem 5.4, 34

problem: prove that if A is symmetric then so is A^+

answer:

Assuming complex matrix then symmetric means $A = A^H$ hence

$$\begin{aligned}
PDQ &= (PDQ)^H \\
&= Q^H D^H P^H \\
&= Q^H D P^H \qquad (1)
\end{aligned}$$

Hence $PDQ = Q^H D P^H \rightarrow DQ = P^{-1}Q^H D P^H \rightarrow D = P^{-1}Q^H D P^H Q^{-1}$

But since $P^H = P^{-1}$ and $Q^H = Q^{-1}$ then the above becomes

$$D = P^H Q^H D P^H Q^H \qquad (2)$$

now

$$A^+ = Q^H D^+ P^H$$

Sub (2) into the above equation we obtain

$$\begin{aligned}
A^{+} &= Q^{H}\left(P^{H}Q^{H}DP^{H}Q^{H}\right)^{+}P^{H}\\
&= Q^{H}\left(\left(P^{H}Q^{H}\right)^{H}D^{+}\left(P^{H}Q^{H}\right)^{H}\right)P^{H}\\
&= Q^{H}\left(\left(QP\right)D^{+}\left(QP\right)\right)P^{H}\\
&= Q^{H}QPD^{+}QPP^{H}
\end{aligned}$$

But $Q^{H}Q=I$ and $PP^{H}=I$

Hence the above becomes

$$A^{+}=PD^{+}Q \tag{3}$$

But

$$\begin{aligned}
\left(A^{+}\right)^{H} &= \left(Q^{H}D^{+}P^{H}\right)^{H}\\
&= PD^{+}Q
\end{aligned} \tag{4}$$

Compare (3) and (4), they are the same.

Hence A^{+} is symmetric.

2.2.2.5 Problem 5.4, 39

problem: prove that eigenvalues of positive semi definite matrix are nonnegative

answer:

positive semi definite means $\vec{x}^{T}A\vec{x}\geq 0$ for all $\vec{x}\neq\vec{0}$

Hence $\vec{x}^{T}A\vec{x}=\vec{x}^{T}\lambda\vec{x}=\lambda\vec{x}^{T}\vec{x}$

But $\vec{x}^{T}\vec{x}=\left\Vert\vec{x}\right\Vert^{2}$

Hence $\vec{x}^{T}A\vec{x}=\lambda\left\Vert\vec{x}\right\Vert^{2}$

We are told the above is ≥ 0. Assume $\vec{x}\neq 0$, then we have $\lambda\times$ the norm, which is positive quantity ≥ 0 , hence this is possible only if λ was zero (for the =0 case) or $\lambda>0$ for the >0 case. It is not possible to have λ negative and multiply it by positive quantity and obtain a positive quantity.

Now Assume $\vec{x}=0$, hence the norm is zero. Hence $A\vec{x}=\vec{0}$ and so eigenvalues is zero. Hence eigenvalues can be either positive or zero. Hence nonnegative

2.2.3 code

This is some code in Matlab

```
%compare the error in A by SVDing it and then rebuild A back

clear all;
A=rand(10000,2);
fprintf('total storage needed for A is =%f MB\n',(8*size(A,1)*size(A,2))/10^6);

[p,d,q]=svd(A,'econ');
econSize=8*(size(p,1)*size(p,2)+size(d,1)*size(d,2)+size(q,1)*size(q,2));
econSize=econSize/10^6; %make it in MB
fprintf('total storage needed for A with economy SVD=%f MB\n',econSize);

fprintf('Max difference between A and its reconstructed version is\n %f\n', ...
         max(max(A-p*d*q')));
```

2.3 Second lecture

2.3.1 Introduction

These are the lecture notes of the second lecture for the course Math 501, Spring 2007 at CSUF. These notes are written by Nasser M. Abbasi (student in the class).

Lecture was given by Dr C.H.Lee, Mathematics dept. CSU Fullerton on 1/24/2007.

Lecture started with walkthrough on using Microsoft powerpoint for presentations.

Note the following correction to last lecture note. For the error term in the 2D Taylor expansion, it is given by

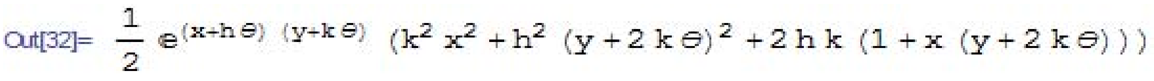
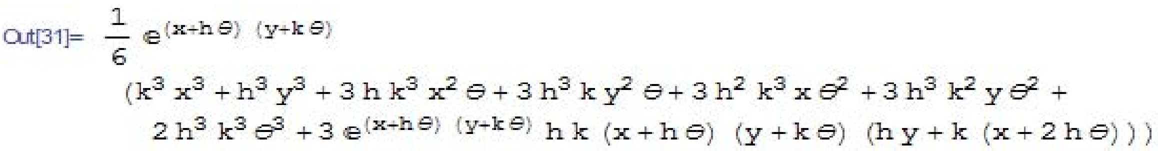
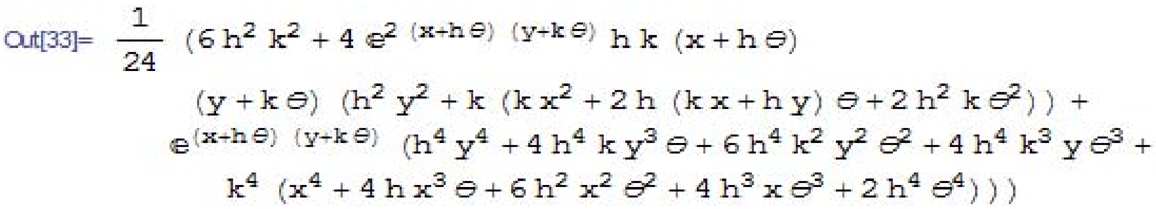
$$E_n\left(x,y\right)=\frac{1}{(n+1)!}\left(h\frac{\partial}{\partial x}+k\frac{\partial}{\partial x}\right)^{n+1}f\left(x+\theta h,y+\theta k\right)$$

Where $0\leq\theta\leq1$

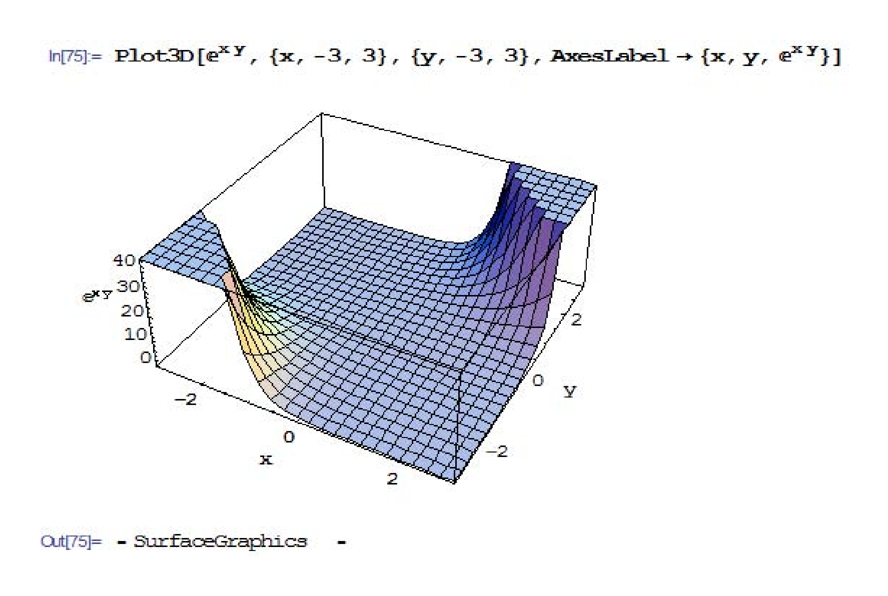
For example, for $n=1$, and for $f\left(x,y\right)=e^{xy}$,the above works out to be

$$E_1\left(x,y\right)=hk+\frac{1}{2}\left(h\left(y+\theta k\right)+k\left(x+\theta h\right)\right)^2e^{(x+\theta h)\left(y+\theta k\right)}$$

I wrote a small function which computes the above error term for any n, here is some of the terms for increasing n

n	E_n
1	 Out[32]= $\frac{1}{2} e^{(x+h\theta)(y+k\theta)} \left(k^2 x^2 + h^2 (y+2k\theta)^2 + 2hk(1+x(y+2k\theta))\right)$
2	 Out[31]= $\frac{1}{6} e^{(x+h\theta)(y+k\theta)} (k^3 x^3 + h^3 y^3 + 3hk^3x^2\theta + 3h^3ky^2\theta + 3h^2k^3x\theta^2 + 3h^3k^2y\theta^2 + 2h^3k^3\theta^3 + 3e^{(x+h\theta)(y+k\theta)} hk(x+h\theta)(y+k\theta)(hy+k(x+2h\theta)))$
3	 Out[33]= $\frac{1}{24} (6h^2k^2 + 4e^{2(x+h\theta)(y+k\theta)} hk(x+h\theta)(y+k\theta)(h^2y^2 + k(kx^2 + 2h(kx+hy)\theta + 2h^2k\theta^2)) + e^{(x+h\theta)(y+k\theta)} (h^4y^4 + 4h^4ky^3\theta + 6h^4k^2y^2\theta^2 + 4h^4k^3y\theta^3 + k^4(x^4 + 4hx^3\theta + 6h^2x^2\theta^2 + 4h^3x\theta^3 + 2h^4\theta^4)))$

To see how the 2D error term behaves as n increases, we can select some values for θ, h, k and evaluate E_n. This is the result for $\theta = 0.5, h = 0.1, k = 0.1$. First, this is a plot of the function e^{xy}

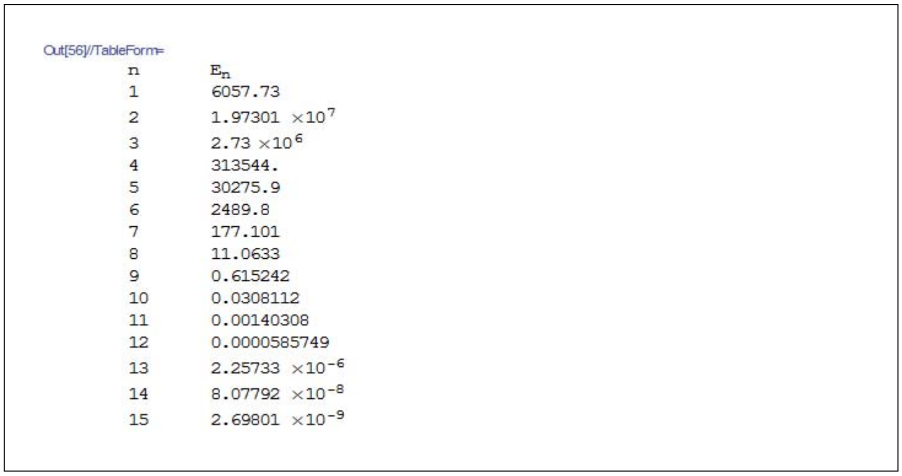


```
In[75]:= Plot3D[e^xy, {x, -3, 3}, {y, -3, 3}, AxesLabel → {x, y, e^xy}]
```

```
Out[75]= - SurfaceGraphics -
```

Figure 2.2: Plot

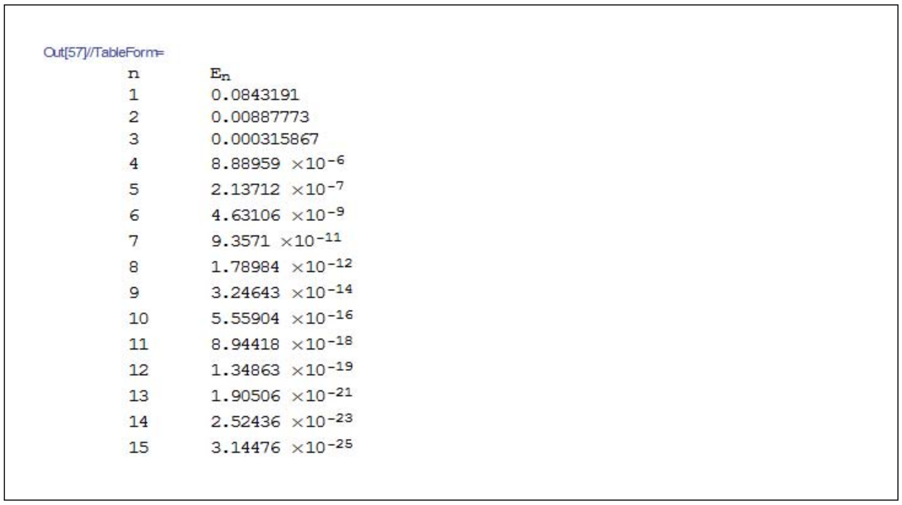
For $(x, y) = (2, 5)$



```
Out[56]//TableForm=
        n       E_n
        1       6057.73
        2       1.97301 ×10^7
        3       2.73 ×10^6
        4       313544.
        5       30275.9
        6       2489.8
        7       177.101
        8       11.0633
        9       0.615242
        10      0.0308112
        11      0.00140308
        12      0.0000585749
        13      2.25733 ×10^-6
        14      8.07792 ×10^-8
        15      2.69801 ×10^-9
```

Figure 2.3: Table

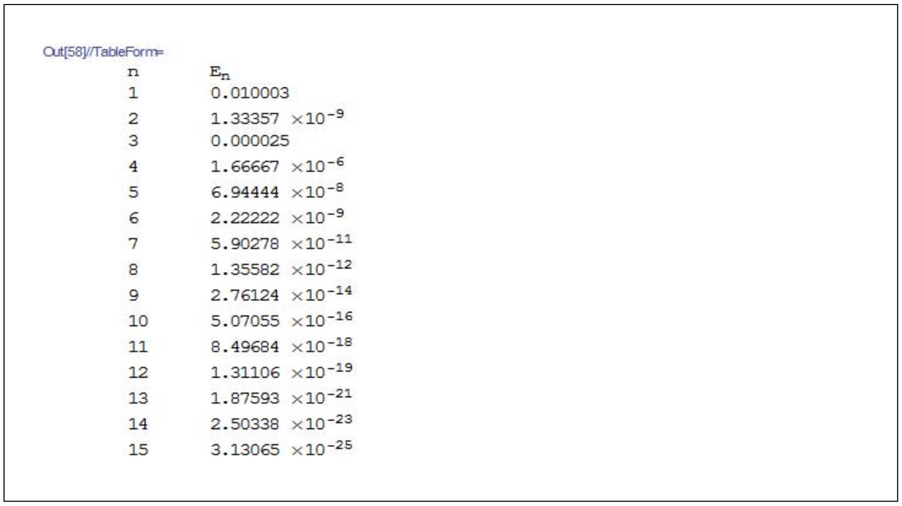
For $\left(x,y\right) = (1,1)$



```
Out[57]//TableForm=
        n       E_n
        1       0.0843191
        2       0.00887773
        3       0.000315867
        4       8.88959 ×10^-6
        5       2.13712 ×10^-7
        6       4.63106 ×10^-9
        7       9.3571 ×10^-11
        8       1.78984 ×10^-12
        9       3.24643 ×10^-14
        10      5.55904 ×10^-16
        11      8.94418 ×10^-18
        12      1.34863 ×10^-19
        13      1.90506 ×10^-21
        14      2.52436 ×10^-23
        15      3.14476 ×10^-25
```

Figure 2.4: Plot

For $\left(x,y\right) = (0,0)$



```
Out[58]//TableForm=
        n      En
        1      0.010003
        2      1.33357 ×10^-9
        3      0.000025
        4      1.66667 ×10^-6
        5      6.94444 ×10^-8
        6      2.22222 ×10^-9
        7      5.90278 ×10^-11
        8      1.35582 ×10^-12
        9      2.76124 ×10^-14
        10     5.07055 ×10^-16
        11     8.49684 ×10^-18
        12     1.31106 ×10^-19
        13     1.87593 ×10^-21
        14     2.50338 ×10^-23
        15     3.13065 ×10^-25
```

Figure 2.5: Plot

Notice the following: If point $\left(x,y\right)$ selected maps to a large function value $f\left(x,y\right)$ then more terms as needed to make the error term smaller. Notice we need at least 4 terms before E_n would continue to decrease as n increases. For example, for $n = 2$, E_2 was actually smaller than E_3 in the examples above. It is only after $n = 4$ than the error term would continue to become smaller as n in increased.

2.3.2 Convergence

Definition: Order of convergence:

Let $[x_n]$ be a sequence, $[x_n]$ is said to converge to L at order P if $\exists$ two positive constants $c < 1$ and $N > 0$ s.t.

$$|x_{n+1} - L| \leq c\,|x_n - L|^P$$

for all $n \geq N$

Note than if we write $E_n \equiv |x_n - L|$, then the above can be written as

$$E_{n+1} \leq c\,(E_n)^P$$

Taking the log, we write

$$\ln E_{n+1} \leq \ln c + P \ln E_n$$

Which can be considered an equation of a line with intercept $\ln c$ and slope P

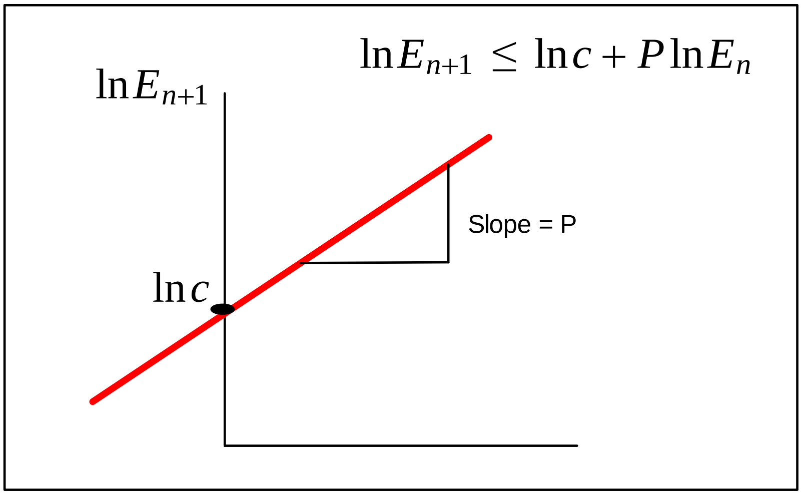


Figure 2.6: Plot

The larger the slope P the faster the sequence $[x_n]$ will converge to L

Example

Suppose $\lim_{n->\infty} [x_n] -> L$ at order 2 (quadratic), then

$$E_{n+1} \leq c \ (E_n)^2$$

For some positive constant c. (Note lecture notes said $c < 1$ here, but text book says c not necessarily less than one for the quadratic case). Now suppose $E_1 = 10^{-1}$, then

$$\begin{aligned}
E_2 &\leq cE_1 \leq c\left(10^{-1}\right)^2 \leq c\,10^{-2}\\
E_3 &\leq c\,(E_2)^2 \leq c\left(c\,10^{-2}\right)^2 \leq c^3 10^{-4}\\
E_4 &\leq c\,(E_3)^2 \leq c\left(c^3 10^{-4}\right)^2 \leq c^7 10^{-8}\\
E_5 &\leq c\left(c^7 10^{-8}\right)^2 \leq c^{15} 10^{-16}
\end{aligned}$$

So after 5 iterations, error became very small. Error went from order 10^{-1} to order 10^{-16}. Compare that to something that converges linearly:

Suppose $\lim_{n->\infty} [x_n] -> L$ at order 1 (Linear), then

$$E_{n+1} \leq c \ E_n$$

For some positive constant $c < 1$ (note for linear case, c must be less than one). Now suppose $E_1 = 10^{-1}$, then

$$\begin{aligned} E_2 &\leq cE_1 \leq c10^{-1} \\ E_3 &\leq cE_2 \leq c\left(c\,10^{-1}\right) \leq c^2 10^{-1} \\ E_4 &\leq cE_3 \leq c\left(c^2 10^{-1}\right) \leq c^3 10^{-1} \\ E_5 &\leq c\left(c^3 10^{-1}\right) \leq c^4 10^{-1} \end{aligned}$$

Compare that to the quadratic case above.

Any $P > 1$ is considered *superlinear*

Example:

Show that $x_{n+1} = \frac{1}{2}x_n + \frac{1}{x_n}$ is a sequences which converges to $L = \sqrt{2}$ quadratically. Use $x_1 = 2$.

We want to show that $E_{n+1} \leq c\left(E_n\right)^2$ for some positive c (note: Lecture notes said also that $c < 1$ here) and for some $n > N$

$$\begin{aligned} E_{n+1} &= \left|x_{n+1} - L\right| \\ &= \frac{1}{2}x_n + \frac{1}{x_n} - \sqrt{2} \\ &= \frac{x_n^2 - \sqrt{2}}{2x_n} - \sqrt{2} \\ &= \frac{x_n^2 - 2\sqrt{2}\,x_n + 2 + 2\sqrt{2}\,x_n}{2x_n} - \sqrt{2} \\ &= \frac{\left(x_n - \sqrt{2}\right)^2 + 2\sqrt{2}\,x_n}{2x_n} - \sqrt{2} \\ &= \frac{\left(x_n - \sqrt{2}\right)^2}{2x_n} + \sqrt{2} - \sqrt{2} \\ &= \frac{\left(x_n - \sqrt{2}\right)^2}{2x_n} \\ &= \frac{\left(E_n\right)^2}{2\left(E_n + \sqrt{2}\right)} \\ &\leq \frac{\left(E_n\right)^2}{2\sqrt{2}} \end{aligned}$$

Hence it converges quadratically, with $c = \frac{1}{2\sqrt{2}}$

Definition: Big O : Let $[x_n]$ and $[\alpha_n]$ be 2 sequences. We say that

$$x_n = O(\alpha_n)$$

if $\exists\, C > 0$ and $N > 0$ s.t. $|x_n| \leq C\,|\alpha_n|$ for all $n \geq N$ (Note: Book only says that C and N are constants, lecture notes says they are also positive).

Example: Given $x_n = \frac{n+1}{n^2}$ and $\alpha_n = \frac{1}{n}$ show that $x_n = O(\alpha_n)$

$$\begin{aligned} x_n &= \frac{n+1}{n^2} \\ &= \frac{1}{n} + \frac{1}{n^2} \\ &\leq \frac{1}{n} + \frac{1}{n} = 2\left(\frac{1}{n}\right) \end{aligned}$$

Hence

$$x_n \leq 2(\alpha_n)$$

Hence $x_n = O(\alpha_n)$

Example: Find a simpler sequence $[\alpha_n]$ s.t. $x_n = O(\alpha_n)$ where $x_n = \frac{\sqrt{n}}{(3n+1)^3}$

$$\begin{aligned} x_n &= \frac{\sqrt{n}}{(3n+1)^3} \\ &\leq \frac{\sqrt{n}}{(3n)^3} \\ &= \frac{1}{9}\frac{n^{\frac{1}{2}}}{n^3} \\ &= \frac{1}{9}n^{-3+\frac{1}{2}} \\ &= \frac{1}{9}n^{-\frac{5}{2}} \end{aligned}$$

Hence $x_n = O\left(\frac{1}{n^{\frac{5}{2}}}\right)$ where here $\alpha_n = \frac{1}{n^{\frac{5}{2}}}$ and $C = \frac{1}{9}$

Definition: small o. Given 2 sequences $[x_n]$ and $[\alpha_n]$ we say $x_n = o(\alpha_n)$ if $\lim_{n\to\infty} \frac{x_n}{\alpha_n} = 0$. i.e. $\exists$ sequence $[\varepsilon_n]$ that converges to zero, and $N > 0$ s.t. $|x_n| \leq |\varepsilon_n|\,|\alpha_n|$ for all $n > N$

The above means that $x_n \to 0$ much faster than α_n does. Graphically, this is illustrated as follows

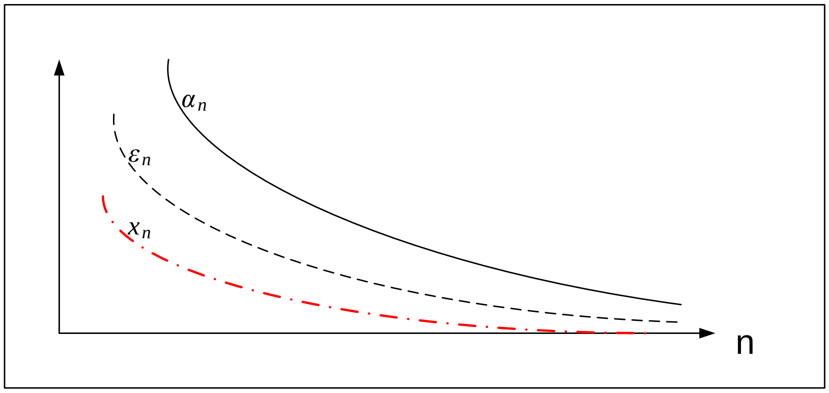


Figure 2.7: Plot

Example: $x_n = \frac{1}{n \ln n}$, hence $x_n = o\left(\frac{1}{n}\right)$ where here $\varepsilon_n = \frac{1}{\ln n}$

Also we can write $x_n = o\left(\frac{1}{\ln n}\right)$ where $\varepsilon_n = \frac{1}{n}$

Example: Show that $e^{-n} = o\left(n^{-k}\right)$ for any $k > 0$

One way is to show that $\lim_{n\to\infty} \frac{x_n}{\alpha_n} = 0$

$$\begin{aligned}\lim_{n\to\infty} \frac{x_n}{\alpha_n} &= \lim_{n\to\infty} \frac{n^k}{e^n}\\ &= \lim_{n\to\infty} \frac{n^k}{1 + n + \frac{n^2}{2} + \frac{n^3}{3!} + \cdots}\\ &= 0\end{aligned}$$

Hence $x_n = o(\alpha_n)$

HW section 1.2, problems 6(b,e), 7(b,c), 10,28,40

Note: Next week we meet in room MH380

2.3.3 Difference equations

Definition: Let us denote by $[x_n] = [x_1, x_2, x_3, \cdots]$ an infinite sequence of numbers. Then E is called the shift operator if $E[x] = [x_2, x_3, x_4, \cdots]$

Some properties of E

1. $Ex_1 = x_2$
2. $Ex_n = x_{n+1}$
3. $E(Ex_1) = x_3$
4. $E^k x_n = x_{n+k}$
5. $E^0 x_n = x_n$

Definition: x_{n+1} is called a recursive sequence if $x_{n+1} = f(x_n, x_{n-1}, \cdots, x_r)$ where $r \leq n$

Example: $x_{n+1} = 2x_n^2 - 7x_{n+1} + 5$. Here 2 initial conditions are needed for this sequence to be defined for all n. In this example the sequence is not linear.

Definition: x_{n+1} is called a *linear* recursive sequence if $x_{n+1} = a_n x_n + a_{n-1}x_{n-1} + \cdots + a_r x_r$ where $r \leq n$. (i.e. x_{n+1} is a linear combination of previous numbers in the sequence)

Given a linear recursive sequence our goal is to find its solution explicitly. i.e. in a non-recursive form.

$$\begin{aligned}
x_{n+1} &= a_n x_n + a_{n-1}x_{n-1} + \cdots + a_r x_r \\
E^{n+1-r}x_r &= a_n E^{n-r}x_r + a_{n-1}E^{n-r-1}x_r + \cdots + a_r E^0 x_r \\
E^{n+1-r}x_r - a_n E^{n-r}x_r - a_{n-1}E^{n-r-1}x_r - \cdots - a_r E^0 x_r &= 0 \\
\left(E^{n+1-r} - a_n E^{n-r} - a_{n-1}E^{n-r-1} - \cdots - a_r E^0\right)x_r &= 0
\end{aligned}$$

Hence since $x_r \neq 0$ we obtain a polynomial in E in degree $n+1-r$

$$E^{n+1-r} - a_n E^{n-r} - a_{n-1}E^{n-r-1} - \cdots - a_r = 0$$

We can now find its roots. Assume the roots are called $\lambda_1, \lambda_2, \cdots, \lambda_{n+1-r}$ then

case 1 all roots λ are distinct (simple roots) then the solution to the difference equation is

$$x_n = k_1\lambda_1^n + k_2\lambda_2^n + \cdots + k_{n+1-r}\lambda_{n+1-r}^n$$

Where $k_1, k_2, \cdots, k_{n+1-r}$ are coefficients to be determined from initial conditions.

Example: Suppose we have linear recursive equation $x_{n+1} = 3x_n - 2x_{n-1}$ with $x_1 = 1, x_2 = 0$ apply the shift operator, we write

$$
\begin{aligned}
E^2 x_{n-1} &= 3E^1 x_{n-1} - 2E^0 x_{n-1} \\
E^2 x_{n-1} - 3E^1 x_{n-1} + 2E^0 x_{n-1} &= 0 \\
\left(E^2 - 3E^1 + 2E^0\right) x_{n-1} &= 0
\end{aligned}
$$

Hence the polynomial is $E^2 - 3E^1 + 2 = 0$ and its roots are given by $(E-2)(E-1) = 0$,hence $\lambda_1 = 2, \lambda_2 = 1$. Simple roots. Hence the solution is given by

$$
\begin{aligned}
x_n &= k_1 \lambda_1^n + k_2 \lambda_2^n \\
&= k_1 2^n + k_2 1^n \\
&= k_1 2^n + k_2
\end{aligned}
$$

Now apply initial conditions to find k_1 and k_2

$n = 1, x_1 = 1 \Rightarrow 1 = k_1 2 + k_2$

$n = 2, x_2 = 0 \Rightarrow 0 = k_1 4 + k_2$

Solving the above 2 equations for k_1, k_2 we obtain $k_1 = -\frac{1}{2}, k_2 = 2$ Hence the solution is

$$
x_n = -\frac{1}{2} 2^n + 2
$$

Case 2: Roots are multiple roots. Let λ_* be k multiple root. Then

$$
\begin{aligned}
x_n &= A_0 \lambda_*^n + A_1 \frac{d}{d\lambda_*}\left(\lambda_*^n\right) + A_2 \frac{d^2}{d\lambda_*^2}\left(\lambda_*^n\right) + \cdots + A_{k-1} \frac{d^{k-1}}{d\lambda_*^{k-1}}\left(\lambda_*^n\right) \\
&= A_0 \lambda_*^n + A_1 n \lambda_*^{n-1} + A_2 n(n-1) \lambda_*^{n-2} + \cdots + A_{k-1} n(n-1)(n-2)\cdots(n-k+2)\lambda_*^{n-k+1}
\end{aligned}
$$

Example: Solve $4x_n = -7x_{n-1} - 2x_{n-2} + x_{n-3}$

The polynomial in E is

$$
\begin{aligned}
4E^3 x_{n-3} + 7E^2 x_{n-3} + 2E^1 x_{n-3} - E^0 x_{n-3} &= 0 \\
4E^3 + 7E^2 + 2E^1 - 1 &= 0 \\
(E+1)^2 (4E-1) &= 0
\end{aligned}
$$

Hence $\lambda_* = -1$ with multiplicity $k = 2$ and $\lambda_1 = \frac{1}{4}$ a simple root.

Solution is

$$
\begin{aligned}
x_n &= \overbrace{\left[k_0\left(\lambda_1\right)^n\right]}^{simple} + \overbrace{\left[A_0\lambda_*^n + A_1\frac{d}{d\lambda_*}\left(\lambda_*^n\right)\right]}^{multiple\ root} \\
&= k_0\left(\frac{1}{4}\right)^n + \left[A_0\left(-1\right)^n + A_1 n\left(\lambda_*^{n-1}\right)\right] \\
&= k_0\left(\frac{1}{4}\right)^n + A_0\left(-1\right)^n + A_1 n\left(-1\right)^{n-1}
\end{aligned}
$$

Where the 3 coefficients k_0, A_0, A_1 can be found from initial conditions.

Next we address stability of these difference equations. For this we need definitions of stable and bounded sequence.

Definition: Bounded: A sequence $[x_n]$ is bounded if $\exists\, c > 0$ and $N > 0$ s.t. $|x_n| \leq c\ \forall\, n \geq N$

Definition: Stability: A difference equation is stable if its solutions are bounded.

Theorem: Let x_n be the solution of the characteristic polynomial of the difference equation. The solution of the difference equation is stable if

1. All simple roots ≤ 1
2. All repeated roots < 1

Proof: Suppose $\lambda's$ are the simple roots of the characteristic polynomial of the difference equation, then

$$
\begin{aligned}
|x_n| &= \left|k_1\lambda_1^n + k_2\lambda_2^n + \cdots + k_{n-r+1}\lambda_{n-r+1}^n\right| \\
&\leq |k_1|\left|\lambda_1^n\right| + |k_2|\left|\lambda_2^n\right| + \cdots + |k_{n-r+1}|\left|\lambda_{n-r+1}^n\right|
\end{aligned}
$$

The above is bounded if each $\lambda \leq 1$

Now assume the roots are multiple roots order k. Then if $|\lambda| < 1$ then

$$\lim_{n\to\infty} n^k\lambda^n = 0$$

Hence $|x_n|$ is bounded. See textbook page 33 for more detailed proof.

Example:

$$x_{n+2} = 3x_{n+1} - 2x_n$$

The characteristic equation is

$$
\begin{aligned}
E^2 - 3E + 2 &= 0 \\
(E-1)(E-2) &= 0
\end{aligned}
$$

Hence $\lambda_1 = 1, \lambda_2 = 2$. Since simple roots, and one root $\lambda_2 > 1$ then NOT stable.

Example:

$$x_{n+2} - 2x_{n+1} + 2x_n = 0$$

The characteristic equation is

$$E^2 - 2E + 2 = 0$$

The roots are $\lambda_1 = 1 - i, \lambda_2 = 1 + i$

Since simple roots, and the size of the root is > 1, then NOT stable (the size of the root is $|\lambda_2| = \sqrt{2}$)

HW, section 1.3 problem 9,11,12,25

2.3.4 Computer arithmetic

2.3.4.1 Decimal system (base 10)

The digits are $0-9$ and each digit it multiplied by power of 10. For example the number 427.325 in base 10 can be written as follows

$$(427.325)_{10} = 4 \times 10^2 + 2 \times 10^1 + 7 \times 10^0 + 3 \times 10^{-1} + 2 \times 10^{-2} + 5 \times 10^{-3}$$

2.3.4.2 Binary system (base 2)

The digits are $0,1$ and each digit is multiplied by powers of 2. For example the number 1001.11101 in base 2 can be written as

$$
\begin{aligned}
(1001.11101)_2 &= 1 \times 2^3 + 0 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 + 1 \times 2^{-1} + 1 \times 2^{-2} + 1 \times 2^{-3} + 0 \times 2^{-4} + 1 \times 2^{-5} \\
&= 8 + 0 + 0 + 1 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + 0 + \frac{1}{32} \\
&= (9.90625)_{10}
\end{aligned}
$$

Example: Calculate $\frac{1}{10}$

In decimal, the answer is 0.1

In Binary (depending on machine accuracy) the answer is

$$\begin{aligned} 1 \div 10 &\simeq (1.000\cdots)_2 \div (1010.000\cdots)_2 \\ &= (0.00011001100110011\cdots)_2 \end{aligned}$$

With 7 digits accuracy machine we get

$$1 \div 10 \simeq 0.10000000\overbrace{1490116}^{error}\cdots$$

Lets do conversion between binary and decimal

Example: Convert 53_{10} to binary

$$\begin{aligned} 53 \div 2 &= 26 \;\; R\,1 \\ 26 \div 2 &= 13 \;\; R\,0 \\ 13 \div 2 &= 6 \;\; R\;\,1 \\ 6 \div 2 &= 3 \;\; R\,0 \\ 3 \div 2 &= 1 \;\; R\,1 \\ 1 \div 2 &= 0 \;\; R\;\,1 \end{aligned}$$

Hence $53_{10} = 110101_2$

Example: Convert 19_{10} to binary

$$\begin{aligned} 19 \div 2 &= 9 \;\; R\,1 \\ 9 \div 2 &= 4 \;\; R\,1 \\ 4 \div 2 &= 2 \;\; R\;\,0 \\ 2 \div 2 &= 1 \;\; R\,0 \\ 1 \div 2 &= 0 \;\; R\,1 \end{aligned}$$

Hence $19_{10} = 10011_2$ **Example:** Convert 0.7_{10} to binary

$$\begin{aligned} 0.7 \times 2 &= 0.4 + 1 \\ 0.4 \times 2 &= 0.8 + 0 \\ 0.8 \times 2 &= 0.6 + 1 \\ 0.6 \times 2 &= 0.2 + 1 \\ 0.2 \times 2 &= 0.4 + 0 \\ 0.4 \times 2 &= 0.8 + 0 \\ &\vdots \\ &\textit{repated} \end{aligned}$$

Hence $0.7_{10} = 0.1\,\overbrace{0110}\ \overbrace{0110}\cdots$

Hence $53.7_{10} = \ 110101.1\,\overbrace{0110}\ \overbrace{0110}\cdots$

2.4 Mathematics Colloquium notes

Mathematics Colloquium notes

Talk given by Dr McMillen,Tyler on 4/25/2007.

Mathematical Problems of Decision Making

California State University, Fullerton.

Spring 2007 semester

Notes taken by Nasser Abbasi.

August 15, 2021

These are my notes taken during talk given by Dr McMillen, Mathematics department, California State University, Fullerton. On April 25, 2007. The subject of the talk was on *Mathematical Problems of Decision Making*

Dr McMillen started by asking the question: ”How to choose between different choices?”, examples given are: run or fight? and asked also: you might have to select between many choices, not just 2.

What are the choice-reaction model?

Need to vary signal to noise ratio and check how people can choose.

If some choices are close to each others, and one choice is distinct one, people tend to select the distinct one. For example, it is easier for people to choose between 2 bars that are drawn at 90 degree to each others, than 2 that are inclined such as they are very close to each others. The first case makes selecting easier since the choices are more distinct from each others.

There is a limit on how many choices people can handle at the same time. The limit seems to be around 7.

Now the talk went into discussing models of decision making:

This is a hard problem. Simplest types of models are only partially understood.

Statistical regularities

Reaction time (RT) effect:

1. Hick’s Law, where $RT \sim \log(N)$ where N is the number of choices
2. Loss avoidance, this means people prefer choices that are far away from each others.
3. The magic number 7.

Stochastic differential equations are used to model decision making process. Mention of Fokker-Planck equation.

Now the talk presented a neural model of decisions making. Where 2 brain neurons are shown each accepting a separate input (with noise added), and there exist what is called an inhibition W factor between these 2 neurons. These neurons are subject also to a decay K factor. This is called *Neural Model of perceptual choices.*

The talk also discussed the effect on the amount of time a person has to make a decision on what decisions they make. When the time to make a decision is limited, it is called the interrogation model.

The talk now discussed what is an optimal method to decide between more than 2 random choices to select from.

Using the above neural model, the best decision is made when the inhibition between the 2 neurons and the decay factor is the same. A model by the name of SPRT (WALD):

Wald's Sequential Probability Ratios, was mentioned in relation to optimally theorem of decision making.

The conclusion of this talk was that a mathematical model of how the brain makes decisions is very complex problem and not well understood, and only a very simple model exist when it comes to making a decision between 2 choices. The optimal way to make a decision is an unsolved mathematical problem.

I found this talk a bit hard to follow. I could not make a clear distinction on how the neural model shown related to the stochastic differential equations presented earlier. I did not understand what does the inhibition factor between neuron mean, and what is the decay factor actually represent? I think the talk was a little advanced for me as I felt I did not completely follow all the points presented. But I did get from this talk that modeling a decision making in humans at the neural level is a very difficult problem, but it was not clear to me why and how this difficulty comes about. Never the less, I did find the talk interesting and informative.

Chapter 3

study notes

Local contents

3.1 Section 1.1

3.1.1 definition of limit of function $f(x)$

We say that the limit to $f(x)$ is L as x gets close to c, if for each number ε we can find another number δ such that $\left|f(x) - L\right| < \varepsilon$ for all x within a distance δ from c.

So if we change ε, may be make it smaller, we need to find another δ, most likely smaller than before also, such that $\left|f(x) - L\right| < \varepsilon$ inside this new interval around c

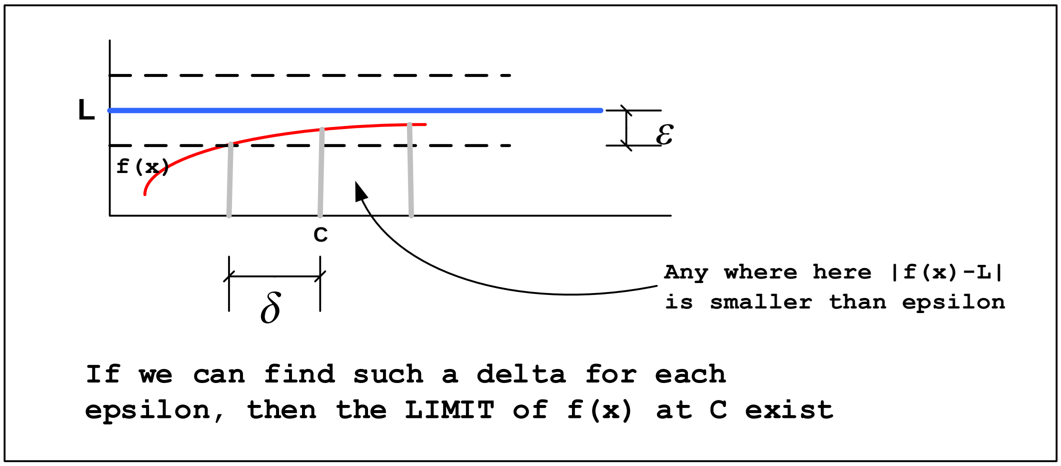

L
f(x)
C
ε
δ
Any where here |f(x)-L| is smaller than epsilon
If we can find such a delta for each epsilon, then the LIMIT of f(x) at C exist

Figure 3.1: Limit

note: We say $\lim_{x\to c} f(x)$ **exist** if $\lim_{x\to c^-} f(x) = \lim_{x\to c^+} f(x) = L$

Example of a function where $\lim_{x->0} f(x)$ does not exist is $f(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ +1 & x > 0 \end{cases}$

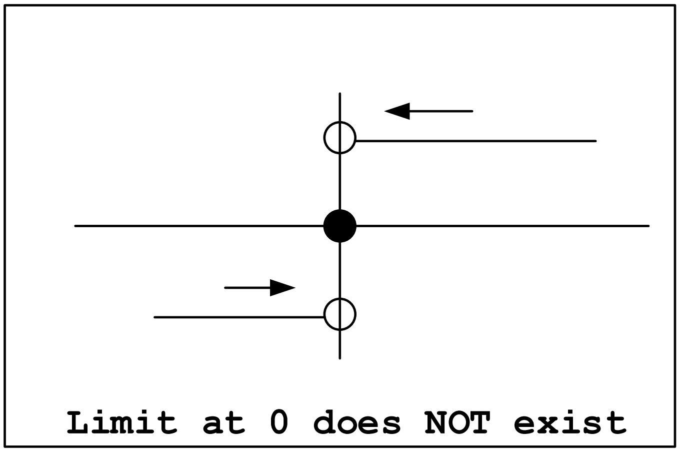


Figure 3.2: example

note: A function can be defined at a point, but not have a limit at that point (as the example above shows)

3.1.2 Definition of continuous function at a point

A function $f(x)$ is continuous at $x = c$ if it is defined at that point, and if $\lim_{x->c} f(x)$ exist and is equal to $f(c)$

Example: of a function that has $\lim_{x->c} f(x)$ exist, but $f(c)$ is not equal to this limit. hence not continues at $x = c$

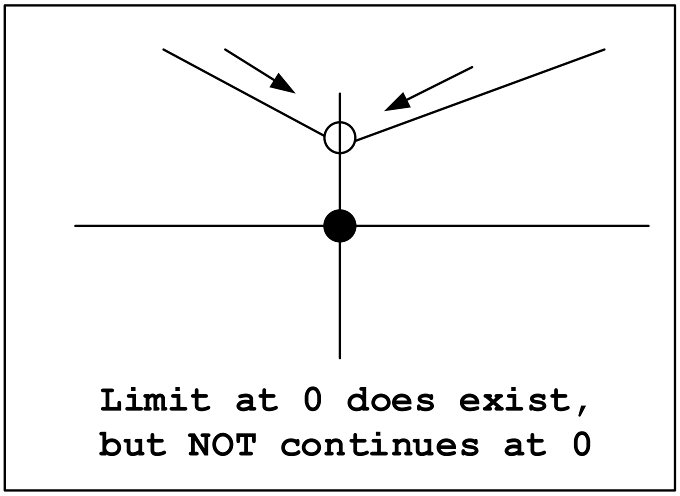


Figure 3.3: example

Example of function where $f(c)$ equal the limit at $x = c$, and $\lim_{x->c} f(x)$ exist, hence continues

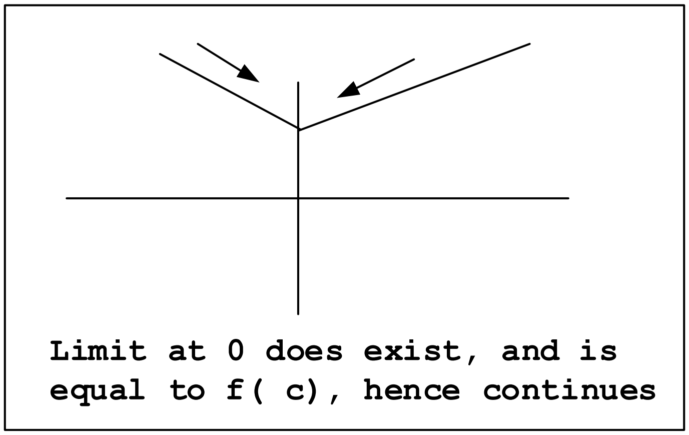

Limit at 0 does exist, and is
equal to f(c), hence continues

Figure 3.4: example

3.1.3 Definition of derivative of function $f(x)$ at c

if $f(x)$ is continues at $x = c$, then

$$f'(c) = \lim_{x \to c} \frac{f(x) - f(c)}{x - c}$$

note: The above $f'(c)$ is defined only if the limit exist and is the same as we approach c from either side.

Conversely, we say that a function $f(x)$ is differentiable at $x = c$ iff $f'(c)$ exist and $f(c)$ is continues.

In other words, $f(x)$ is differentiable at $x = c$ iff $\lim_{x \to c^-} \frac{f(x)-f(c)}{x-c} = \lim_{x \to c^+} \frac{f(x)-f(c)}{x-c} = f'(c)$

note: It is possible for a function to be continues at c but not be differentiable there if the above limit is not the same as we approach c from either side.

Example, $f(x) = |x|$

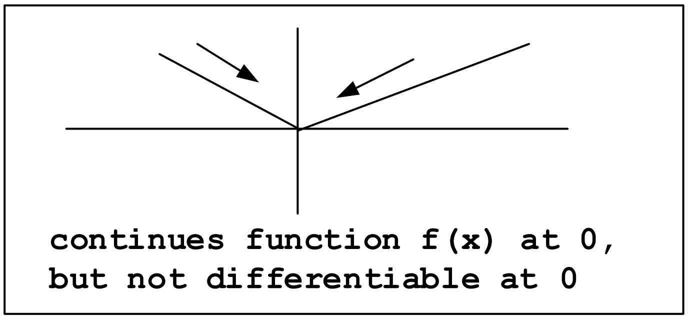
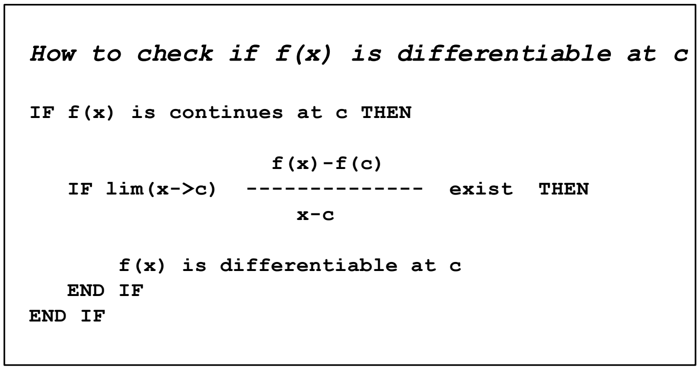


Figure 3.5: example



```
How to check if f(x) is differentiable at c

IF f(x) is continues at c THEN

                   f(x)-f(c)
   IF lim(x->c)  --------------  exist  THEN
                     x-c

       f(x) is differentiable at c
   END IF
END IF
```

Figure 3.6: example

3.1.4 Intermediate value theorem

on interval $[a, b]$, a continues function assumes all values between $f(a)$ and $f(b)$

3.2 Taylor expansion with Lagrange remainder

On the real line, if we have a function $f(x)$, and we wish to know the value of this function at a point $x = b$ given the value of $f(x)$ and its derivatives at another point say $x = a$, then we write

$$f(b) = f(a) + (b-a)f'(a) + \frac{(b-a)^2 f''(a)}{2!} + \cdots$$

Now suppose we want to find the value of the function at arbitrary point x given the value of $f(x)$ and its derivative at another point say $x = a$, then we replace b by x above and write

$$f(x) = f(a) + (x-a)f'(a) + \frac{(x-a)^2 f''(a)}{2!} + \cdots + R_n$$

Where

$$R_n = \frac{(x-a)^{n+1}}{(n+1)!} f^{(n+1)}(\xi)$$

Where ξ is some point between x and a

If $x - a = h$, we can write the above as

$$\tilde{f}(x) = f(a) + hf'(a) + \frac{h^2 f''(a)}{2!} + \frac{h^3 f''(a)}{3!} + \cdots + \frac{h^{n+1}}{(n+1)!} f^{(n+1)}(\xi)$$

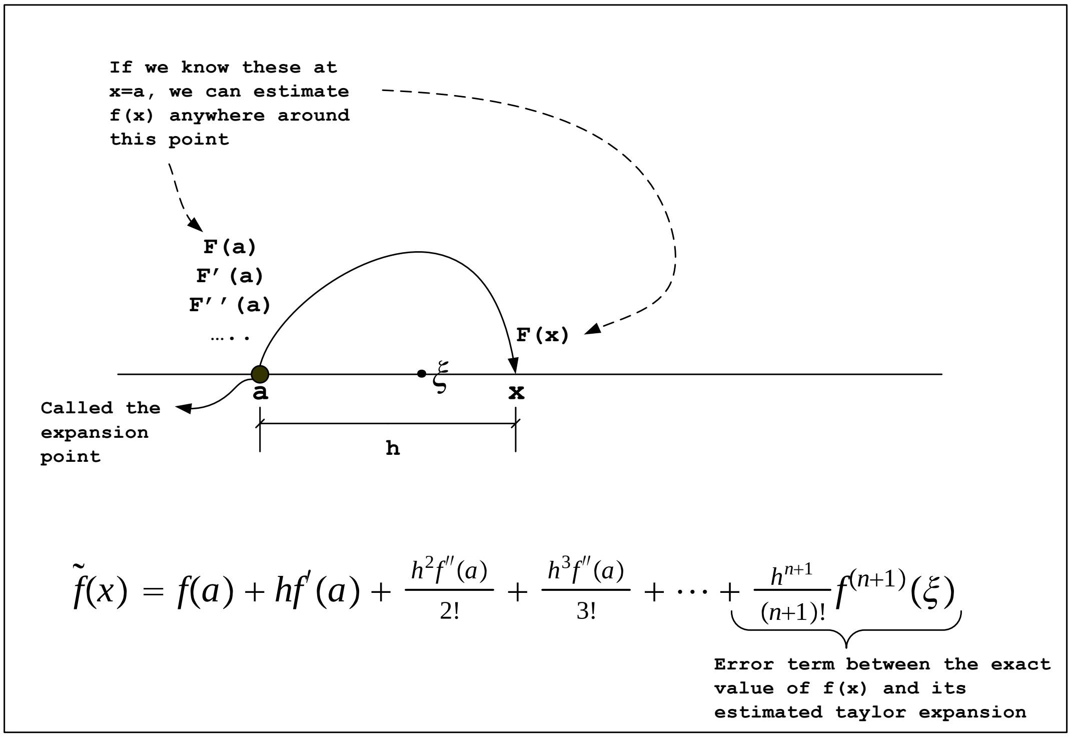

If we know these at x=a, we can estimate f(x) anywhere around this point
F(a)
F' (a)
F'' (a)
.....
F(x)
ξ
a
x
Called the expansion point
h
$\tilde{f}(x) = f(a) + hf'(a) + \frac{h^2 f''(a)}{2!} + \frac{h^3 f''(a)}{3!} + \cdots + \underbrace{\frac{h^{n+1}}{(n+1)!} f^{(n+1)}(\xi)}$
Error term between the exact value of f(x) and its estimated taylor expansion

Figure 3.7: example

note: If the point of expansion is zero, Taylor series is called maclaurin series.

$$\tilde{f}(x) = f(a) + xf'(0) + \frac{x^2 f''(0)}{2!} + \frac{x^3 f''(0)}{3!} + \cdots + \frac{x^{n+1}}{(n+1)!} f^{(n+1)}(\xi)$$

Why do we use Taylor series for? To express a function as a series. This can allow one to more easily manipulate it. Also, if the function is non-linear, by expressing it in series, and dropping low order non-linear terms (h must be very small to have good approximation), then we have linearized a non-linear function in the vicinity of a point of expansion. Hence around the point of expansion, we can approximate the non-linear function by its linear Taylor series terms for the purpose of doing further linear system analysis (as it is easier to work with linear functions than non-linear ones).

3.2.1 Finding Error in Taylor series approximation

Things to know: How to find how many terms in Taylor series to approximate some given function to some accuracy?

Idea of solution: Express E_n, this is the error term, or the remainder. Make $|E_n| < \epsilon$ where ϵ is the accuracy needed. Find smallest n which makes this true

Example: How many terms needed to find ln(2) to accuracy of $\epsilon = 10^{-8}$?

Expand $\ln(x)$ at $x = 1$, hence $h = 2 - 1 = 1$

$$\begin{aligned}\ln(x) &= (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 + \cdots + E_n \\ \ln(2) &= 1 - \frac{1}{2} + \frac{1}{3} + \cdots + E_n\end{aligned}$$

We want $|E_n| < 10^{-8}$, but $E_n = \frac{1}{n+1} < 10^{-8} \Rightarrow n \geq 10^8$, hence at least 100 million terms would be needed to computer ln(2) using Taylor series with accuracy of 10^{-8}

3.2.2 Mean value theorem

if $f(x)$ is continues on $[a, b]$, and if $f'(x)$ exist on the open interval (a, b) then there exists a point ξ between b, a s.t.

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

3.2.3 Rolle's theorem

if $f(x)$ is continuous on $[a,b]$ and if $f'(x)$ exist on (a,b) and if $f(a) = f(b)$ then $f'(\xi) = 0$ for some point in (a,b)

3.3 Section 1.2 Order of convergence, linear, superlinear, quadratic

3.3.1 convergent sequence, limit definition

definition: A sequence of numbers x_n converges to limit x^*, if for every positive number ϵ there exist a number r such that $|x_n - x^*| < \epsilon$ for all $n > r$

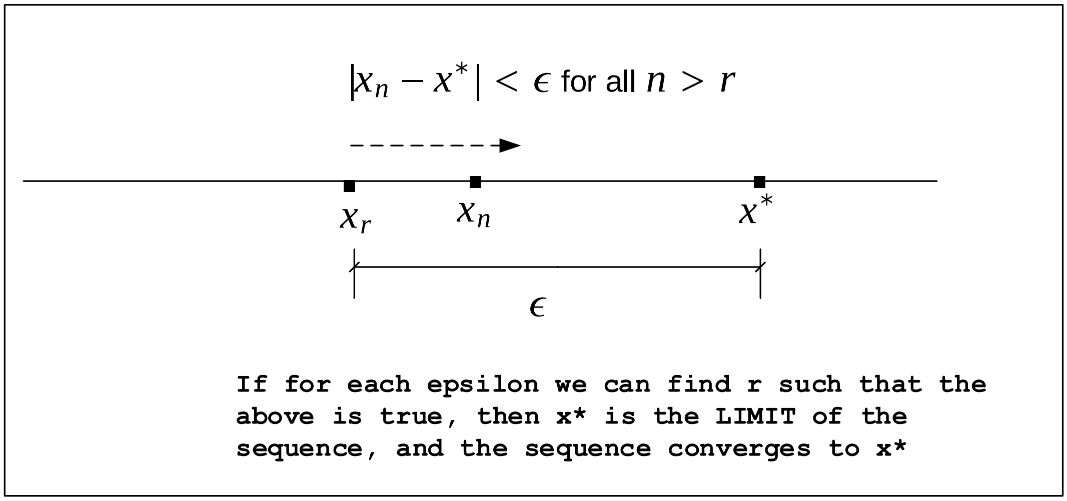

$|x_n - x^*| < \epsilon$ for all $n > r$
x_r
x_n
x^*
ϵ
If for each epsilon we can find r such that the above is true, then x* is the LIMIT of the sequence, and the sequence converges to x*

Figure 3.8: example

note: Hence to show a sequence converges to x^*, all what we have to do is find r such that for any given ϵ, we get $|x_n - x^*| < \epsilon$ whenever $n > r$

Example: to show that $x_n = \frac{n+1}{n}$ converges to 1, need to show that there exist a number r such that $\left|\frac{n+1}{n} - 1\right| < \epsilon$ whenever $n > r$. Rewrite we have $\left|\frac{1}{n}\right| < \epsilon$, hence we see that $r = \epsilon^{-1}$, because whenever $n > r$, then $\left|\frac{1}{n}\right| < \epsilon$. For example, assume $\epsilon = 0.1$, then $r = 10$, and whenever $n > 10$, we have $\frac{1}{n} < 0.1$

It is clear that $\lim_{n->\infty} \frac{n+1}{n} = 1$.

3.3.2 order of convergence

The order of the convergence of a sequence is the largest number q s.t. $\lim_{n\to\infty}\frac{|x_{n+1}-x^*|}{|x_n-x^*|^q}$ exist

This is the same as writing $\lim_{n\to\infty}\frac{|e_{n+1}|}{|e_n|^q}$ where e_n is the error at x_n

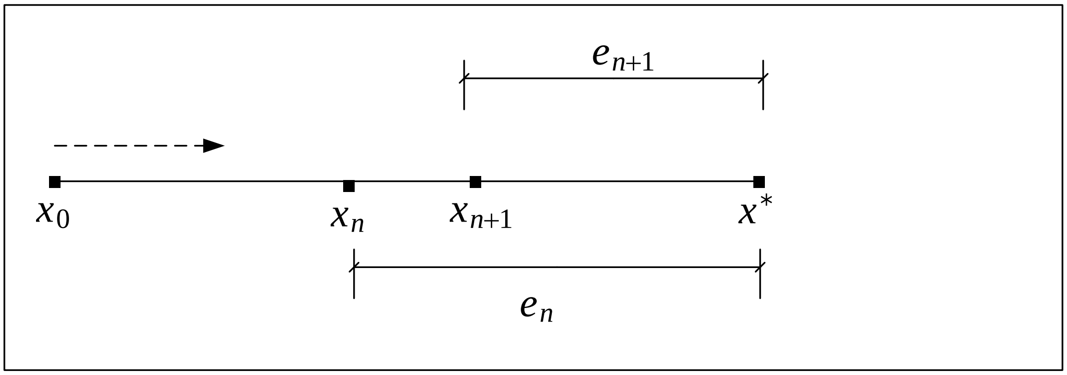


Figure 3.9: example

3.3.2.1 Linear

The rate of convergence is **linear** if we can find constant $c < 1$ and integer N s.t.

$$|e_{n+1}| \leq c\ |e_n| \qquad n \geq N$$

3.3.2.2 superlinear

The rate of convergence is superlinear if we can find sequence ε_n tending to zero and integer N s.t.

$$|e_{n+1}| \leq \varepsilon_n\ |e_n| \qquad n \geq N$$

question: I do not understand this definition Can I say it as the linear, but an exponent $1 < \alpha < 2$, and write

The rate of convergence is **superlinear** if we can find constant $c < 1$ and integer N s.t.

$$|e_{n+1}| \leq c\ |e_n|^{\alpha} \qquad n \geq N$$

3.3.2.3 quadratic

The rate of convergence is **quadratic** if we can find constant c NOT Necessarily less than 1, and integer N s.t.

$$|e_{n+1}| \leq c \ |e_n|^2 \qquad\qquad n \geq N$$

idea: To show that a sequence converges quadratically to some limit, start with the expression for e_{n+1} and manipulate it to show that that is it $\leq$ some constant $\times e_n$

3.3.3 Big O and little o

These are means by which to compare 2 sequences to each others.

def: big O: One sequence x_n is bounded by a linear scaled version of a second sequence

we say that $x_n = O\alpha_n$ if there is a constant C and N, s.t. $x_n \leq C\alpha_n$ for all $n > N$

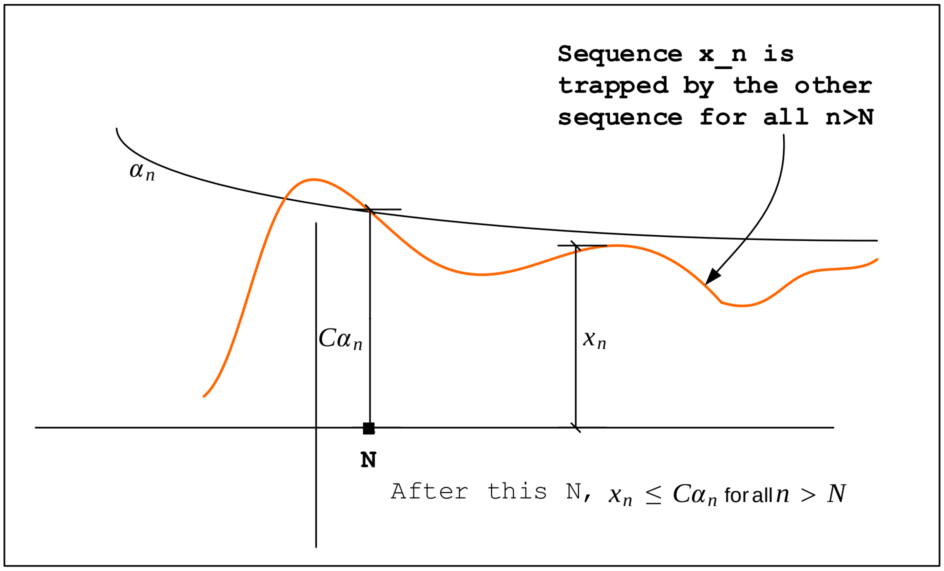


Figure 3.10: example

How to find if $x_n = O(\alpha_n)$? start with x_n expression and manipulate it so that it has only terms that contain α_n with some multipliers (the constant).

Or, easier, just look to see if x_n is bigger or smaller than α_n, if it is bigger, then it goes to zero AFTER α_n, hence use BIG O. if it is smaller, then it goes to zero BEFORE α_n, hence us little o.

def: little o: we say $x_n = o(\alpha_n)$ if $\lim_{n\to\infty} \frac{x_n}{\alpha_n} = 0$

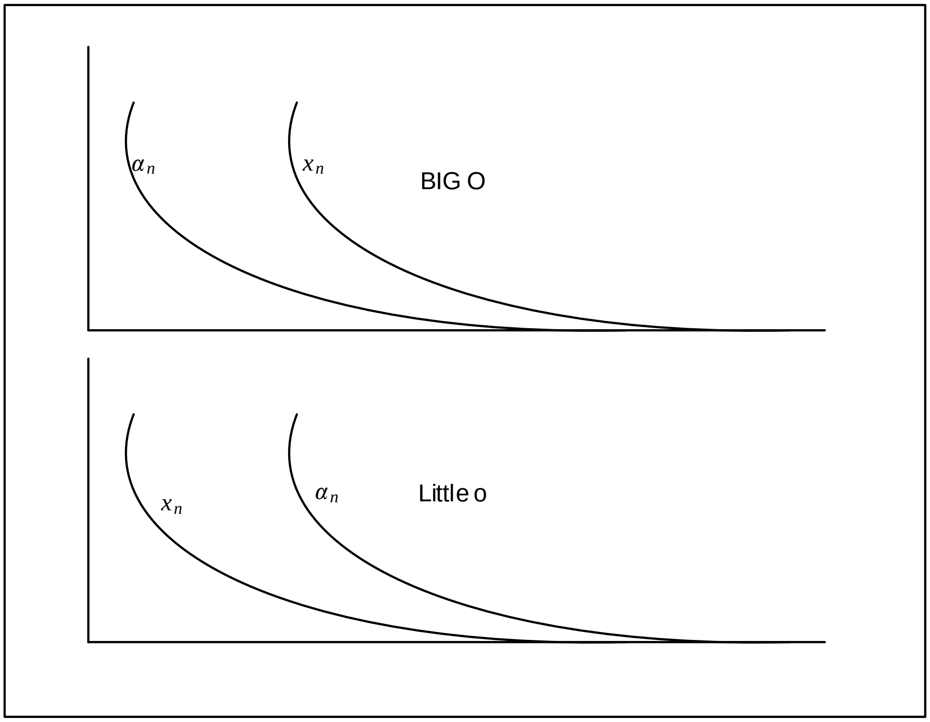


Figure 3.11: example

To test the above, given $x_n = \frac{n+1}{n^2}$ and $\alpha_n = \frac{1}{n}$, what is the relation between them? we see that $x_n = \frac{1}{n} + \frac{1}{n^2}$, so it is BIGGER than α_n, so it goes to zero AFTER. Hence $x_n = O\alpha_n$

given $x_n = \frac{1}{n \ln n}$, we see that this is SMALLER than α_n, hence it will go to zero BEFORE α_n, hence $x_n = o(\alpha_n)$

question: verify I can do this reasoning all the time.

3.4 Section 1.3. difference equations, characteristic polynomial, simple and repeated roots, analytic solution and stability

Given a linear recursive equation such as $x_{n+1} = f\left(x_n, x_{n-1,} \cdots\right)$, the goal is to find an non-recursive solution for x_n

To do that, we introduce the shift operator E, rewrite the recursive equation in terms of E, we get a polynomial in E which we solve. and depending on how the root come out, we get a solution for x_n, which is called the explicit solution.

Notice that the explicit solution gives the value of x_n right away as a function of n. No recursion is needed to find x for some specific n, hence one can get numerical problems with the recursive solution due to cancellation errors, while the explicit solution will not show this problem. It is always better to use the explicit solution.

3.4.1 Example simple roots, no repeated roots

$$\begin{aligned} x_{n+1} &= 3x_n - 2x_{n-1} \\ E^2 x_{n-1} &= 3E x_{n-1} - 2E^0 x_{n-1} \\ \left(E^2 - 3E + 2\right) x_{n-1} &= 0 \end{aligned}$$

Solve $P(E) = 0$, we get roots $\lambda_1 = 2, \lambda_2 = 1$, hence simple distinct roots. Hence explicit solution is

$$\begin{aligned} x_n &= A_1 \lambda_1^n + A_2 \lambda_2^n \\ &= A_1 2^n + A_2 \end{aligned}$$

Now A_1 and A_2 can be found from initial conditions

3.4.2 Example simple roots, repeated roots

$$4x_n + 7x_{n-1} + 2x_{n-2} - x_{n-3} = 0$$

$$P(E) = 4E^3 x_{n-3} + 7E^2 x_{n-3} + 2E x_{n-3} - E^0 x_{n-3} = 0$$

Hence roots are found from $(E+1)^2 (4E-1) = 0$, so $\lambda_1 = \lambda_2 = -1$ repeated 2 times, and $\lambda_3 = \frac{1}{4}$

hence solution is

$$x_n = \left(A_1 \lambda_1^n + A_2 n \lambda_2^{n-1}\right) + A_3 \lambda_3^n$$

so it a root is repeated k times, we write

$$x_n = \left(A_1 \lambda_1^n + A_2 n \lambda_2^{n-1} + A_3 n (n-1) \lambda_3^{n-2} + \cdots + A_k n (n-1)(n-2) \cdots \left(\lambda_k^{n-k+1}\right)\right) + A_{k+1} \lambda_{k+1}^n$$

so here we have

$$x_n = A_1 (-1)^n + A_2 n (-1)^{n-1} + A_3 \left(\frac{1}{4}\right)^n$$

and use I.C. to find the coefficients.

3.4.3 Bounded sequence

sequence $x_n = [x_1, x_2, \cdots]$ is bounded if there is a constant c s.t. $|x_n| \leq c$ for all n. i.e. $\sup_n |x_n| < \infty$

3.4.3.1 Stable solution for the difference equation $P(E)x = 0$

$P(E)x = 0$ is stable if all its solutions are stable.

Theorem: polynomial p satisfying $P(0) \neq 0$ the difference equation $P(E)x = 0$ is stable iff all simple roots of $p(E) = 0$ are ≤ 1 and all repeated roots are < 1

So to find if a recursive equation is stable, find the roots of $P(E) = 0$ and check as per above

3.5 Section 3.1 Bisection

After n iterations, c_n , which is $\left(\frac{a_n+b_n}{2}\right)$ is at a distance from root given by

$$|c_n - r| = \frac{b_0 - a_0}{2^{n+1}}$$

3.6 Section 3.2 Newton root finding

Understanding Newton method

Start from the line equation. You remember we normally write it as

$y = c + mx$ where c is the y-axis intercept and m is the slope. But this equation always implicitly assumed that the slope is taken at $x_0 = 0$, and the intercept is also at $x_0 = 0$, hence the above can be written as

$$y = f(x) = f(x_0) + f'(x_0)(x - x_0) \tag{1}$$

The above equation is the same as $y = c + mx$ when $x_0 = 0$

So from (1)

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) \\ \frac{f(x) - f(x_0)}{f'(x_0)} &= (x - x_0) \end{aligned}$$

Now instead of writing x and x_0 we write x_{n+1} and x_n, so the above becomes

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

That is it. Now for x_{n+1}, $f(x_{n+1}) = 0$, and that is the whole idea. Replace $f(x_{n+1})$ by zero in the above we obtain

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

Which is Newton method.

Theorem: Let f'' be C^2 and let r be a simple zero. There is a neighborhood of r and a constant C s.t. if Newton method is started in that neighborhood it will converge to r according to $|x_{n+1} - r| \leq C\,|x_n - 1|^2$

Proof of quadratic convergence order

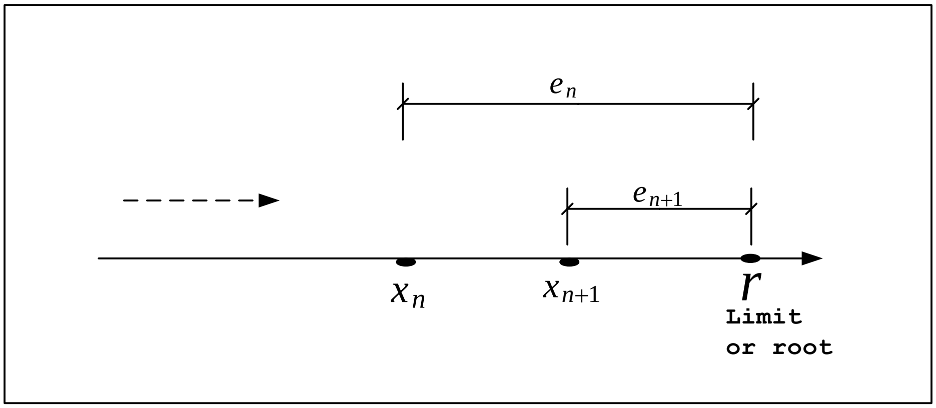


Figure 3.12: quadratic convergence

From diagram we see that

$$e_{n+1} = r - x_{n+1} \tag{1}$$

But from definition of Newton root finding

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

substituting (2) into (1) gives

$$
\begin{aligned}
e_{n+1} &= r - \left(x_n - \frac{f(x_n)}{f'(x_n)} \right) \\
&= \overbrace{r - x_n}^{e_n} + \frac{f(x_n)}{f'(x_n)} \\
&= e_n + \frac{f(x_n)}{f'(x_n)} \\
&= \frac{e_n f'(x_n) + f(x_n)}{f'(x_n)} \qquad (3)
\end{aligned}
$$

Now evaluate $f(x)$ at r by expanding it using Taylor around x_n

$$
f(r) = f(x_n) + h f'(x_n) + \frac{h^2 f''(\xi)}{2!}
$$

But h is the distance between r and x_n, which is e_n, hence the above becomes

$$
f(r) = f(x_n) + e_n f'(x_n) + \frac{e_n^2 f''(\xi)}{2!}
$$

But $f(r) = 0$ since this is a root finding, and that is our goal. Hence the above becomes

$$
\begin{aligned}
0 &= f(x_n) + e_n f'(x_n) + \frac{e_n^2 f''(\xi)}{2!} \\
f(x_n) + e_n f'(x_n) &= -\frac{f''(\xi)}{2!} e_n^2 \qquad (4)
\end{aligned}
$$

Substituting (4) into numerator of (3) gives

$$
\begin{aligned}
e_{n+1} &= \frac{-\frac{f''(\xi)}{2!} e_n^2}{f'(x_n)} \\
&= -\frac{f''(\xi)}{2 f'(x_n)} e_n^2
\end{aligned}
$$

but $\frac{f''(\xi)}{f'(x_n)} \approx \frac{f''(r)}{f'(r)} = k$ (some constant), hence above becomes

$$
e_{n+1} = k_1 e_n^2
$$

where k_1 is some constant as well.

Hence we can find a constant C such that $\boxed{e_{n+1} \leq Ce_n^2}$ where C is any constant smaller than k_1

The above proofes that Newton method converges quadratically.

3.6.1 Definition of convex function

A function $f(x)$ is convex if $f''(x) \geq 0$ for all x.

Mathworld has this definition

"A convex function is a continuous function whose value at the midpoint of every interval in its domain does not exceed the arithmetic mean of its values at the ends of the interval." diagram below from Mathworld

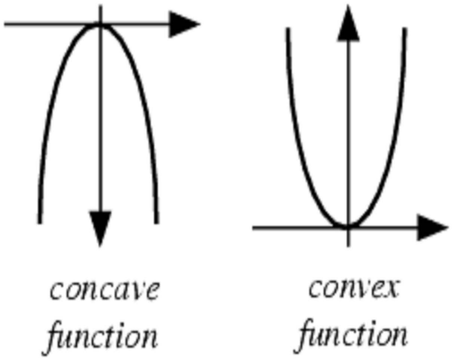


Figure 3.13: Mathworld

How to use Newton method to find $\sqrt{R}$?

Let $x = \sqrt{R}$, hence $x^2 - R = 0$ is $f(x)$ to use with Newton method. This leads to $x_{n+1} = x_n - \frac{x_n^2 - R}{2x_n} = \frac{1}{2}\left(x_n + \frac{R}{x_n}\right)$

3.6.2 Newton method to solve set of equations

Writing

$$
\begin{aligned}
\begin{bmatrix} x_{n+1} \\ y_{n+1} \end{bmatrix} &= \begin{bmatrix} x_n \\ y_n \end{bmatrix} - F\left(x_n\right) J^{-1}\left(x_n\right) \\
&= \begin{bmatrix} x_n \\ y_n \end{bmatrix} - \begin{bmatrix} F_1\left(x_n, y_n\right) \\ F_2\left(x_n, y_n\right) \end{bmatrix} \begin{bmatrix} \frac{\partial F_1\left(x_n, y_n\right)}{x_n} & \frac{\partial F_1\left(x_n, y_n\right)}{y_n} \\ \frac{\partial F_2\left(x_n, y_n\right)}{x_n} & \frac{\partial F_2\left(x_n, y_n\right)}{x_n} \end{bmatrix}
\end{aligned}
$$

3.7 Section 7.1 Numerical differentiation and Richardson extrapolation

Some points to know

1. If a function $f(x)$ is known at n points, and we also know that the function is a polynomial of *at most* $n-1$ degree, then we can find the polynomial exactly by solving n equations and finding the $c_0, c_1, \cdots, c_n$ coefficients. Hence no need to do numerical differentiation, we can do analytical differentiation.

2. Remember this for Taylor: $f\left(x+h\right) = f(x) + hf'(x) + \frac{h^2}{2} f''\left(\xi\right)$ for this to be valid, $f(x), f'\left(x\right)$ have to be continuous in the CLOSED interval between x and h while $f''(x)$ need to exist on the OPEN interval.

$f\left(\sqrt{2}+h\right) - \frac{df}{dx}@\left(\sqrt{2}\right)$

$f\left(\sqrt{2}\right)$

$f'\left(\sqrt{2}\right) = \frac{f\left(\sqrt{2}+h\right)-f\left(\sqrt{2}\right)}{h}$

Chapter 4

HWs

Local contents

4.1 lookup table

#	date	section/problems	my solution	score
1		1.1: 10,16,24,32 1.2: 6(b,e),7(b,c),10,28,40 1.3: 9,11,12,25		18/20
2		2.2: 9,12,16,21,2.3: 2,4,6,7, 3.1: 2,14,15,16 3.2: 9,15,16,17,19,22,23,32		20/20
3		3.3: 4,5,6, 3.4: 4,5,10,12,13,29,40		18/20
4		3.5: 1,2,3,5,6,10, 4.1: 15,16,17,18, 4.2: 1,5,13,27,30,33,39,47		20/20, 17/20
5	march 2	4.2: 1,5,13,27,30,33,39,47		26/30
6		4.3: 1 (b),(e), 30,31,39,43,45		35/35
7		4.4: 7(a),(c), 21, 37, 40 (a),(c), 4.5: 2,5,8,12,22,24		18/20, 15/15
8	4/4/07	4.6: 2,14,16,17, 4.7: 1,2,6	Including computer assignment on iterative solvers) Richardson, Jacobi, Gauss-Seidel, SOR, Steepest descent	25/25
9	4/24/07	4.7: 9, and computer assignment on finding eigenvalues using power,InversePower,Shifted power, Shifted inverse power		20/20
10	5/3/07	5.3: 2,3,14,16,20,29,30,37		10/20
11	5/8/07	6.2: 13,22,26,27,37, 6.3: 4,9,12,23		
12	5/15/07			100/100

4.2 HW 1

Local contents

4.2.1 Section 1.1, Problem 10

Problem: Prove or disprove this assertion: if f is differentiable at x, then for any $\alpha \neq 1$

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x+\alpha h)}{h - \alpha h}$$

Solution:

Since $f'(x)$ exists, then expanding $f(x+h)$ and $f(x+\alpha h)$ in Taylor series results in

$$\begin{aligned} f(x+h) &= f(x) + f'(x)h + \cdots \text{(Higher Order Terms involving } h^m \text{ where } m \geq 2) \\ f(x+\alpha h) &= f(x) + (\alpha h) f'(x) + \cdots \text{(Higher Order Terms involving } (\alpha h)^m \text{ where } m \geq 2) \end{aligned}$$

From first equation above we write

$$f(x) = f(x+h) - hf'(x) - \text{(Higher Order Terms involving } h^m \text{where } m \geq 2) \qquad (1)$$

And from the second equation we write

$$f(x) = f(x+\alpha h) - (\alpha h) f'(x) - \text{(Higher Order Terms involving } (\alpha h)^m \text{ where } m \geq 2) \quad (2)$$

equating equations (1)-(2)=0 we obtain

$$\begin{aligned} \left[f(x+h) - hf'(x) - O(h^m)\right] - \left[f(x+\alpha h) - (\alpha h) f'(x) - O\ (\alpha h)^m\right] &= 0 \\ f'(x)[\alpha h - h] + f(x+h) - f(x+\alpha h) - O\ (h^m) + O(\alpha h)^m &= 0 \end{aligned}$$

Keep $f'(x)$ on one side, and move everything to the other side results in

$$f'(x) = \frac{f(x+\alpha h) - f(x+h)}{(\alpha h - h)} + \frac{\left(O\ (h^m) - O(\alpha h)^m\right)}{(\alpha h - h)}$$

As h goes to zero the above reduces to

$$f'(x) = \lim_{h\to 0} \frac{f(x+\alpha h) - f(x+h)}{(\alpha h - h)}$$

rearrange the sign results in

$$f'(x) = \lim_{h\to 0} \frac{f(x+h) - f(x+\alpha h)}{(h - \alpha h)}$$

4.2.2 Section 1.1, Problem 16

Problem: If the series for $\ln x$ is truncated after the term involving $(x-1)^{1000}$ and is then used to compute $\ln 2$, what bound on the error can be give?

Answer:

Assume $\ln x$ has a power series expansion around x_0, we write, from definition of power series

$$\ln(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + a_3(x-x_0)^3 + \cdots + a_n(x-x_0)^n + \cdots \tag{1}$$

When $x_0 = 1$ we get

$$\ln(x) = a_0 + a_1(x-1) + a_2(x-1)^2 + a_3(x-1)^3 + \cdots + a_n(x-1)^n + \cdots \tag{2}$$

At $x = 1$ we obtain $a_0 = 0$ since $\ln(1) = 0$

Differentiate (2)

$$\frac{1}{x} = a_1 + 2a_2(x-1) + 3a_3(x-1)^2 + \cdots + na_n(x-1)^{n-1} + \cdots \tag{3}$$

At $x = 1$ we obtain $a_1 = 1$

Differentiate (3)

$$\frac{-1}{x^2} = 2a_2 + (3\times 2)a_3(x-1) + \cdots + n(n-1)a_n(x-1)^{n-2} + \cdots \tag{4}$$

At $x = 1$ we obtain $-1 = 2a_2 \rightarrow a_2 = \frac{-1}{2}$

Differentiate (4)

$$\frac{2}{x^3} = (3\times 2)a_3 + \cdots + n(n-1)(n-2)a_n(x-1)^{n-3} + \cdots \tag{5}$$

at $x = 1$ we obtain $a_3 = \frac{1}{3}$

continue as above, we obtain the power series for $\ln(x)$ as

$$\ln(x) = (x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \frac{1}{4}(x-1)^4 + \cdots + \frac{(-1)^{n+1}}{n}(x-1)^n + \cdots$$

Notice that for the above to converge, we need to have $(x-1) \le 1$, or $x \le 2$

Now if the series is truncated after $(x-1)^{1000}$, hence $n = 1000$, and the maximum error will be the $(n+1)$ term.

Hence

$$\begin{aligned} E &\le \left| \frac{(-1)^{1002}}{1001}(x-1)^{1001} \right| \\ &\le \frac{(x-1)^{1001}}{1001} \end{aligned}$$

which for $x = 2$

$$\begin{aligned} E &\le \frac{(2-1)^{1001}}{1001} \\ &\le \frac{1}{1001} \\ &\le 9.99 \times 10^{-4} \end{aligned}$$

4.2.3 Section 1.1, Problem 24

Problem: For small values of x, how good is the approximation for $\cos x \approx 1 - \frac{1}{2}x^2$? for what range of values will this approximation give correct results rounded to 3 decimal places?

Answer:

Expand $\cos(x)$ in power series, we write

$$\cos(x) = a_0 + a_1(x-x_0) + a_2(x-x_0)^2 + a_3(x-x_0)^3 + \cdots + a_n(x-x_0)^n + \cdots$$

expand at $x_0 = 0$

$$\cos(x) = a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \cdots + a_n x^n + \cdots$$

At $x = 0 \rightarrow a_0 = 1$, Differentiate the above

$$-\sin(x) = a_1 + 2a_2 x + 3a_3 x^2 + \cdots + n a_n x^{n-1} + \cdots$$

at $x = 0 \to a_1 = 0$, Differentiate the above

$$-\cos(x) = 2a_2 + (3\times 2)\,a_3x + \cdots + n\,(n-1)\,a_nx^{n-2} + \cdots$$

at $x = 0 \to a_2 = -\frac{1}{2}$, continue as above, we obtain the series for $\cos(x)$ as

$$\cos(x) = 1 - \frac{1}{2}x^2 + \frac{1}{4!}x^4 - \cdots + \frac{(-1)^{n+1}}{n!}x^n + \cdots$$

Hence if we truncate the series at $\cos(x) \approx 1 - \frac{1}{2}x^2$, then the maximum error will be bounded by

$$E \leq \frac{1}{4!}x^4$$

Since we want the error to be correct to 3 decimal places, then we write

$$E < 0.001$$

Hence

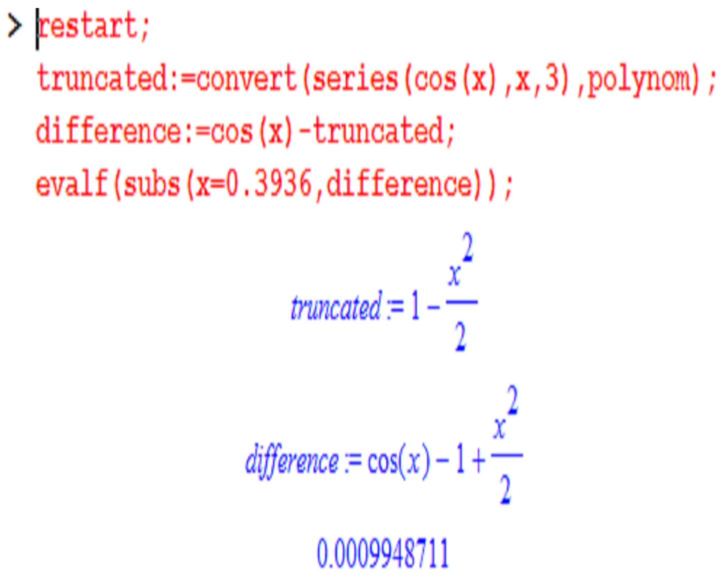
$$\begin{aligned} x^4 &< 4!(0.001) \\ &< 0.024 \end{aligned}$$

Hence

$$x < (0.024)^{\frac{1}{4}} = 0.393\,60$$

So for $x < 0.393\,60$ radians (about 22.552^0), the approximation $\cos(x) \approx 1 - \frac{1}{2}x^2$ give correct results to 3 decimal places.

A small code to verify:



```
> restart;
  truncated:=convert(series(cos(x),x,3),polynom);
  difference:=cos(x)-truncated;
  evalf(subs(x=0.3936,difference));
```

$$truncated := 1 - \frac{x^2}{2}$$

$$difference := \cos(x) - 1 + \frac{x^2}{2}$$

```
0.0009948711
```

Figure 4.1: code

4.2.4 Section 1.1, Problem 32

Problem: First develop the function $\sqrt{x}$ in a series of powers of $(x-1)$ and then use it to approximate $\sqrt{0.99999\,99995}$ to 10 decimal places.

Solution:

$$\sqrt{x} = a_0 + a_1 (x - x_0) + a_2 (x - x_0)^2 + a_3 (x - x_0)^3 + \cdots + a_n (x - x_0)^n + \cdots$$

Expand at $x_0 = 1$

$$\sqrt{x} = a_0 + a_1 (x-1) + a_2 (x-1)^2 + a_3 (x-1)^3 + \cdots + a_n (x-1)^n + \cdots$$

at $x = 1 \rightarrow a_0 = 1$, differentiating the above we obtain

$$\frac{1}{2\sqrt{x}} = a_1 + 2a_2 (x-1) + 3a_3 (x-1)^2 + \cdots + na_n (x-1)^{n-1} + \cdots$$

at $x = 1 \rightarrow a_1 = \frac{1}{2}$, differentiate the above we obtain

$$\frac{-1}{4(x)^{3/2}} = 2a_2 + (3 \times 2)\, a_3 (x-1) + \cdots + n(n-1)\, a_n (x-1)^{n-2} + \cdots$$

at $x = 1 \rightarrow a_2 = -\frac{1}{4\times 2} = -\frac{1}{8}$, differentiate the above we obtain

$$\frac{3}{8(x)^{5/2}} = (3\times 2)\,a_3 + \cdots + n\,(n-1)\,(n-2)\,a_n\,(x-1)^{n-3} + \cdots$$

at $x=1 \rightarrow a_3 = \frac{3}{8(3\times 2)} = \frac{1}{16}$ differentiating the above gives

$$-\frac{15}{16(x)^{7/2}} = (4\times 3\times 2)\,a_4 + \cdots + n\,(n-1)\,(n-2)\,(n-3)\,a_n\,(x-1)^{n-4} + \cdots$$

at $x=1 \rightarrow a_4 = -\frac{15}{16(4\times 3\times 2)} = -\frac{5}{128}$

Hence the series is

$$\sqrt{x} = 1 + \frac{1}{2}(x-1) - \frac{1}{8}(x-1)^2 + \frac{1}{16}(x-1)^3 - \frac{5}{128}(x-1)^4 + \cdots$$

Note: For convergence we require $|x| \leq 1$

We want accuracy to 10 decimal places. Since

$$\sqrt{0.99999\,99995} = 0.9999\,999974$$

Then the series, using 2 terms gives

$$\sqrt{0.99999\,99995} \approx \left(1 + \frac{1}{2}(x-1)\right)_{x=0.9999999995} = 1 + \frac{1}{2}(0.9999999995 - 1) = 0.9999\,999975$$

hence 2 terms are only needed. hence $n = 1$

4.2.5 Section 1.2, problem 6(b,e)

Problem: For the pair (x_n, α_n), is it true that $x_n = O(\alpha_n)$ as $n \rightarrow \infty$?

b) $x_n = 5n^2 + 9n^3 + 1, \alpha_n = 1$

e) $x_n = \sqrt{n+3}, \alpha_n = \frac{1}{n}$

Solution:

b) Assume that $5n^2 + 9n^3 + 1 \leq C(\alpha_n)$ hence $5n^2 + 9n^3 + 1 \leq C$, but since $n > 1$ and keeps increasing, then no matter how large a C we select, $5n^2 + 9n^3$ will eventually become larger than any constant C we choose when $n > N$ for sufficiently large N.

Hence there is no such C, hence the answer is $\boxed{\text{NOT TRUE}}$.

e) we see that $lim_{n\rightarrow\infty}x_n = \infty$, however $\lim_{n\rightarrow\infty}\frac{1}{n} = 0$, hence it is not possible to find C s.t. $\sqrt{n+3} \leq C\frac{1}{n}$ for any N. Hence the answer is $\boxed{\text{NOT TRUE}}$.

4.2.6 Section 1.2, problem 7(b,c)

Problem: Choose the correct assertion (in each, $n \rightarrow \infty$)

b) $\frac{n+1}{\sqrt{n}} = o(1)$

c)$\frac{1}{\ln n} = O\left(\frac{1}{n}\right)$

Solution:

b) $x_n = \frac{n+1}{\sqrt{n}}, \alpha_n = 1$.

$$\begin{aligned}
\lim_{n\rightarrow\infty}\left(\frac{x_n}{\alpha_n}\right) &= \lim_{n\rightarrow\infty}\left(\frac{n+1}{\sqrt{n}}\right)\\
&= \lim_{n\rightarrow\infty}\left(\frac{n+1}{\sqrt{n}}\right)\\
&= \lim_{n\rightarrow\infty}\left(\frac{n}{\sqrt{n}}\right) + \lim_{n\rightarrow\infty}\left(\frac{1}{\sqrt{n}}\right)\\
&= \lim_{n\rightarrow\infty}\left(\sqrt{n}\right) + 0\\
&\neq 0
\end{aligned}$$

Since the limit as $n \rightarrow \infty$ is not zero, hence the assertion is $\boxed{\text{FALSE}}$

c)$x_n = \frac{1}{\ln n}, \alpha_n = \frac{1}{n}$. Since $\ln n$ grows less rapidly than n then $\frac{1}{\ln n}$grows more rapidly than $\frac{1}{n}$, Hence it is not possible to find some constant C s.t. $\frac{1}{\ln n} \leq C\frac{1}{n}$, hence assertion is $\boxed{\text{FALSE}}$

4.2.7 Section 1.2, problem 10

Problem: Show that these assertions are not true:

a) $e^x - 1 = O\left(x^2\right)$ as $x \rightarrow 0$

b) $x^{-2} = O(\cot x)$ as $x \rightarrow 0$

c) $\cot x = o\left(x^{-1}\right)$ as $x \rightarrow 0$

Answer:

a) We need to

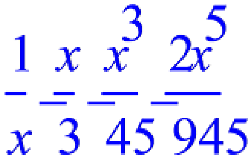
$$\begin{aligned}
x_n &= e^x - 1\\
&= \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right) - 1\\
&= x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots
\end{aligned}$$

and $\alpha_n = x^2$

As $x \rightarrow 0$ the term x will become larger than x^2, hence near $x = 0$, $x_n > \alpha_n$ since near $x = 0$ $x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots > x^2$

Therefore it is not possible to find a constant C such that $x_n \leq C\alpha_n$ near $x = 0$ since for any constant C we select, no matter how small, we can find x closer to zero such that $x_n > C\alpha_n$, Hence assertion is $\boxed{\text{not true}}$.

b) The power series for cot(x) is (Using CAS:).



$$\frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \frac{2x^5}{945}$$

Figure 4.2: code

Here we have $x_n = x^{-2}$ and $\alpha_n = \frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \cdots$

As $x \rightarrow 0$, then $\alpha_n \rightarrow \frac{1}{x}$

But $\frac{1}{x}$ will grow less rapidly than $\frac{1}{x^2}$would as $x \rightarrow 0$, hence it is not possible to find a constant C such that $x_n \leq C\alpha_n$ near $x = 0$ since for any constant C we select, no matter how small, can find x closer to zero such that $x_n > C\alpha_n$, Hence assertion is $\boxed{\text{not true}}$.

c) $x_n = \cot(x) = \frac{1}{x} - \frac{x}{3} - \frac{x^3}{45} - \cdots, \alpha_n = \frac{1}{x}$, hence

$$\begin{aligned}
\lim_{x\to 0}\frac{x_n}{\alpha_n} &= \lim_{x\to 0}\frac{\frac{1}{x}-\frac{x}{3}-\frac{x^3}{45}-\cdots}{\frac{1}{x}}\\
&= \lim_{x\to 0}\frac{\frac{1}{x}\left(1-\frac{x^2}{3}-\frac{x^4}{45}-\cdots\right)}{\frac{1}{x}}\\
&= \lim_{x\to 0}\left(1-\frac{x^2}{3}-\frac{x^4}{45}-\cdots\right)\\
&= 1
\end{aligned}$$

Since the limit does not go to zero, hence the assertion is $\boxed{\text{not true}}$.

4.2.8 Section 1.2, problem 28

Problem: Prove that $x_n = x + o(1)$ iff $\lim_{n\to\infty} x_n = x$

Solution:

Let $X_n = x_n - x, \alpha_n = 1$

Forward direction proof:

$$\begin{aligned}
\lim_{n\to\infty} X_n &= \lim_{n\to\infty}(x_n - x)\\
&= \left(\lim_{n\to\infty} x_n\right) - \left(\lim_{n\to\infty} x\right)\\
&= \left(\lim_{n\to\infty} x_n\right) - x
\end{aligned}$$

If $\lim_{n\to\infty} x_n = x$, then the above become $x - x = 0$

Hence $\lim_{n\to\infty}\frac{X_n}{\alpha_n} = \frac{0}{1} = 0$, hence $X_n = o(1)$ or $x_n - x = o(1)$ or $x_n = x + o(1)$

Now proof in the reverse direction. Assume that $\lim_{n\to\infty} x_n \neq x$, we need to show that this implies $x_n \neq x + o\,(1)$

If $\lim_{n\to\infty} x_n \neq x$, then we can say that $\lim_{n\to\infty} x_n = \beta$, where $\beta \neq x$, hence $\lim_{n\to\infty} X_n = \beta - x$

Hence

$$\begin{aligned}
\lim_{n\to\infty}\frac{X_n}{\alpha_n} &= \lim_{n\to\infty}\frac{\beta - x}{\alpha_n}\\
&= \frac{\beta - x}{1}\\
&= \beta - x
\end{aligned}$$

But since $\beta \neq x$, then this limit does not go to zero. Hence $x_n \neq x + o(1)$. This complete the proof.

4.2.9 Section 1.2, problem 40

Problem: Prove: If $\alpha_n \to 0$, $x_n = O(\alpha_n)$, and $y_n = O(\alpha_n)$ then $x_n y_n = o(\alpha_n)$

Answer: Since $x_n = O(\alpha_n)$ then $x_n \leq C_1(\alpha_n)$, and since $y_n = O(\alpha_n)$ then $y_n \leq C_2(\alpha_n)$, where C_1, C_2 are positive constants.

Hence

$$\begin{aligned} x_n y_n &\leq C_1 C_2 (\alpha_n) \\ &\leq C(\alpha_n) \end{aligned}$$

Where $C = C_1 C_2$

But $x_n y_n \leq C(\alpha_n)$ means that $x_n y_n$ is bounded above by α_n.

But we are told next that $\lim_{n\to\infty} \alpha_n = 0$, hence this means that the sequence $x_n y_n$ will reach zero before the sequence α_n. But this is the same as saying that $x_n y_n = o(\alpha_n)$

4.2.10 Section 1.3, problem 9

Problem: Prove that if L_1 and L_2 are linear combinations of powers of E and if $L_1 x = 0$, then $L_1 L_2 x = 0$

Answer: Let $L_1 = a_1 E^{n_1} + a_2 E^{n_2} + \cdots$ and $L_2 = b_1 E^{m_1} + b_2 E^{m_2} + \cdots$

Then

$$
\begin{aligned}
L_1L_2x &= \left(a_1E^{n_1} + a_2E^{n_2} + \cdots\right)\left(b_1E^{m_1} + b_2E^{m_2} + \cdots\right)x \\
&= \left(a_1E^{n_1} + a_2E^{n_2} + \cdots\right)\left(b_1E^{m_1}x + b_2E^{m_2}x + \cdots\right) \\
\\
&= \left(a_1E^{n_1}\left(b_1E^{m_1}x\right) + a_2E^{n_2}\left(b_1E^{m_1}x\right) + \cdots\right) \\
&+ \left(a_1E^{n_1}\left(b_2E^{m_2}x\right) + a_2E^{n_2}\left(b_2E^{m_2}x\right) + \cdots\right) \\
&+ \cdots \\
&= \left(b_1a_1E^{n_1}\left(E^{m_1}x\right) + b_1a_2E^{n_2}\left(E^{m_1}x\right) + \cdots\right) \\
&+ \left(b_2a_1E^{n_1}\left(E^{m_2}x\right) + b_2a_2E^{n_2}\left(E^{m_2}x\right) + \cdots\right) \\
&+ \cdots \\
&= \left(b_1a_1E^{m_1}\left(E^{n_1}x\right) + b_1a_2E^{m_1}\left(E^{n_2}x\right) + \cdots\right) \\
&+ \left(b_2a_1E^{m_2}\left(E^{n_1}x\right) + b_2a_2E^{m_2}\left(E^{n_2}x\right) + \cdots\right) \\
&+ \cdots \\
&= \left(b_1E^{m_1} + b_2E^{m_2} + \cdots\right)\left(a_1E^{n_1}x + a_2E^{n_2}x + \cdots\right) \\
&= L_2L_1x \\
&= L_2(0) \\
&= 0
\end{aligned}
$$

4.2.11 Section 1.3, Problem 11

Problem: Give bases consisting of real sequences for each solution space.

a) $\left(4E^0 - 3E^2 + E^3\right)x = 0$

b) $\left(3E^0 - 2E + E^2\right)x = 0$

c) $\left(2E^6 - 9E^5 + 12E^4 - 4E^3\right)x = 0$

d) $\left(\pi E^2 - \sqrt{2}\,E + E^0 \log 2\right)x = 0$

Solution:

a) Characteristic equation is $\lambda^3 - 3\lambda^2 + 4 = 0$, or $(\lambda+1)(\lambda-2)^2 = 0$, hence the roots are $\lambda = -1$, and $\lambda = 2$ or multiplicity 2.

i.e. $\lambda_1 = -1, \lambda_2 = 2, \lambda_3 = 2$

Hence first solution $x_1(n)$ associated with $\lambda_1 = -1$ is $x_1(n) = \lambda_1^n = -1^n$

the second solution $x_2(n)$ associated with $\lambda_2 = -2$ is $x_2(n) = \lambda_2^n = 2^n$

the third solution $x_3(n)$ associated with $\lambda_3 = -2$ is $x_3(n) = \frac{dx_2(n)}{d\lambda} = n\lambda_2^{n-1} = n2^{n-1}$

Hence now we can write some terms of the above 3 basis solutions are follows

$$
\begin{aligned}
x_1(n) &= \left[\lambda_1^1, \lambda_1^2, \lambda_1^3, \cdots\right]\\
&= \left[-1, -1^2, -1^3, \cdots\right]\\
&= [-1, 1, -1, 1, \cdots]
\end{aligned}
$$

$$
\begin{aligned}
x_2(n) &= \left[\lambda_2^1, \lambda_2^2, \lambda_2^3, \cdots\right]\\
&= \left[2^1, 2^2, 2^3, \cdots\right]\\
&= [2, 4, 8, 16, 32, \cdots]
\end{aligned}
$$

$$
\begin{aligned}
x_3(n) &= \left[\lambda_2^0, 2\lambda_2^1, 3\lambda_2^2, 4\lambda_2^3, \cdots\right]\\
&= \left[\left(2^0\right), 2\left(2^1\right), 3\left(2^2\right), 4\left(2^3\right), 5\left(2^4\right), \cdots\right]\\
&= [1, 4, 12, 32, 80, \cdots]
\end{aligned}
$$

Hence the basis are

$$
\begin{aligned}
&[-1, 1, -1, 1, \cdots]\\
&[2, 4, 8, 16, 32, \cdots]\\
&[1, 4, 12, 32, 80, \cdots]
\end{aligned}
$$

b)$\left(3E^0 - 2E + E^2\right) x = 0$

Characteristic equation is $\lambda^2 - 2\lambda + 3 = 0$, The roots are

$$
\begin{aligned}
\lambda_1 &= 1 + \sqrt{2}\, i\\
\lambda_2 &= 1 - \sqrt{2}\, i
\end{aligned}
$$

Hence first solution $x_1(n)$ associated with $\lambda_1 = 1 + \sqrt{2}\, i$ is $x_1(n) = \lambda_1^n = \left(1 + \sqrt{2}\, i\right)^n$

the second solution $x_2(n)$associated with $\lambda_2 = 1 - \sqrt{2}\, i$ is $x_2(n) = \lambda_2^n = \left(1 - \sqrt{2}\, i\right)^n$

Hence

$$\begin{aligned} x_1(n) &= \lambda_1^n \\ &= \left[\left(1+\sqrt{2}\,i\right)^1, \left(1+\sqrt{2}\,i\right)^2, \left(1+\sqrt{2}\,i\right)^3, \cdots\right] \\ &= \left[\left(1+\sqrt{2}\,i\right), \left(-1+2i\sqrt{2}\right), \left(-5+i\sqrt{2}\right), \left(-7-4i\sqrt{2}\right), \cdots\right] \end{aligned}$$

$$\begin{aligned} x_2(n) &= \left[\left(1-\sqrt{2}\,i\right)^1, \left(1-\sqrt{2}\,i\right)^2, \left(1-\sqrt{2}\,i\right)^3, \cdots\right] \\ &= \left[\left(1-\sqrt{2}\,i\right), \left(-1-2i\sqrt{2}\right), \left(-5-i\sqrt{2}\right), \left(-7+4i\sqrt{2}\right) \cdots\right] \end{aligned}$$

Notice that the 2 basis are conjugate to each others in each term in the sequence.

c)$\left(2E^6 - 9E^5 + 12E^4 - 4E^3\right) x = 0$

Characteristic equation is $2\lambda^6 - 9\lambda^5 + 12\lambda^4 - 4\lambda^3 = 0$

Factoring we obtain λ

$$^3 (2\lambda - 1)(\lambda - 2)^2 = 0$$

hence the solutions are

$\lambda = 0$ with multiplicity 3, $\lambda = \frac{1}{2}, \lambda = 2$ with multiplicity 2.

Hence Solutions associated with $\lambda = 0$ are

$x_1(n) = \lambda^n, x_2(n) = n\lambda^{n-1}, x_3(n) = n(n-1)\lambda^{n-2}$

Hence $x_1(n) = [0, 0, 0, \cdots]$, and x_2 and x_3 are also the null sequence.

Solution associated with $\lambda = \frac{1}{2}$ is $x_4(n) = \lambda^n = \left(\frac{1}{2}\right)^n = \left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \cdots\right]$

Solutions associated with $\lambda = 2$ are $x_5(n) = \lambda^n = 2^n = [2, 4, 8, 16, \cdots]$

and $x_6(n) = \frac{dx_5}{d\lambda} = n\lambda^{n-1} = n2^{n-1} = \left[1, 2(2), 3\left(2^2\right), 4\left(2^3\right), \cdots\right] = [1, 4, 12, 32, \cdots]$

Hence the basis are

$$\begin{aligned} &[0, 0, 0, \cdots] \\ &\left[\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \cdots\right] \\ &[2, 4, 8, 16, \cdots] \\ &[1, 4, 12, 32, \cdots] \end{aligned}$$

d)$\left(\pi E^2 - \sqrt{2}\,E + E^0 \log 2\right) x$

Characteristic equation is

$$\begin{aligned}\pi\lambda^2 - \sqrt{2}\,\lambda + \log 2 &= 0\\ \lambda^2 - \frac{\sqrt{2}}{\pi}\lambda + \frac{\log 2}{\pi} &= 0\\ \lambda^2 - 0.450\,16\;\lambda + 0.220\,64 &= 0\end{aligned}$$

$$\begin{aligned}\lambda &= \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{0.450\,16 \pm \sqrt{0.450\,16^2 - 4 \times 0.220\,64}}{2}\\ &= \frac{0.450\,16 \pm \sqrt{-0.679\,92}}{2}\\ &= \frac{0.450\,16 \pm i0.824\,57}{2}\end{aligned}$$

Hence $\lambda_1 = 0.225079 + i0.41228$ and $\lambda_2 = 0.225079 - i0.41228$

Hence

$$\begin{aligned}x_1 &= \lambda_1^n\\ &= (0.225079 + i0.41228)^n\end{aligned}$$

and

$$\begin{aligned}x_2 &= \lambda_2^n\\ &= (0.225079 - i0.41228)^n\end{aligned}$$

4.2.12 Section 1.3, Problem 12

Problem: Prove that if P is a polynomial with real coefficients and if $z \equiv [z_1, z_2, z_3, \cdots]$ is a complex solution of $p(E)z = 0$, then the conjugate of z, the real part of z and the imaginary part of z are also solutions.

Solution:

$$P(E)z = 0$$

Take conjugate of both sides

$$\begin{aligned}\overline{P(E)}\;\overline{z} &= \overline{0}\\ \overline{P(E)}\;\overline{z} &= 0\end{aligned}$$

But

$$P(E) = a_0E^0 + a_1E^1 + a_2E^2 + \cdots$$

and all the $a's$ are real, hence $\overline{P(E)} = P(E)$, then

$$P(E)\bar{z} = 0 \tag{1}$$

Now take the real part of $P(E)z = 0$ we get

$$\begin{aligned} \operatorname{Re}(P(E)z) &= \operatorname{Re}(0) \\ \operatorname{Re}(P(E))\operatorname{Re}(z) &= 0 \end{aligned}$$

But

$$\operatorname{Re}(P(E)) = P(E)$$

Hence

$$P(E)\operatorname{Re}(z) = 0 \tag{2}$$

For the last part, let

$$z = \operatorname{Re}(z) + i\ \operatorname{Im}(z)$$

Then $P(E)z = 0$ can be written as

$$\begin{aligned} P(E)\{\operatorname{Re}(z) + i\ \operatorname{Im}(z)\} &= 0 \\ P(E)\operatorname{Re}(z) + i\,P(E)\operatorname{Im}(z) &= 0 \end{aligned}$$

But from (2) we see that $P(E)\operatorname{Re}(z) = 0$, hence the above becomes

$$i\,P(E)\operatorname{Im}(z) = 0$$

Hence

$$P(E)\operatorname{Im}(z) = 0$$

4.2.13 Section 1.3 problem 25

Problem: Determine if the difference equation $x_n = x_{n-1} + x_{n-2}$

Solution: Using the shift operator, we write $E^2 x_{n-2} = E x_{n-1} + E^0 x_{n-2}$

Hence

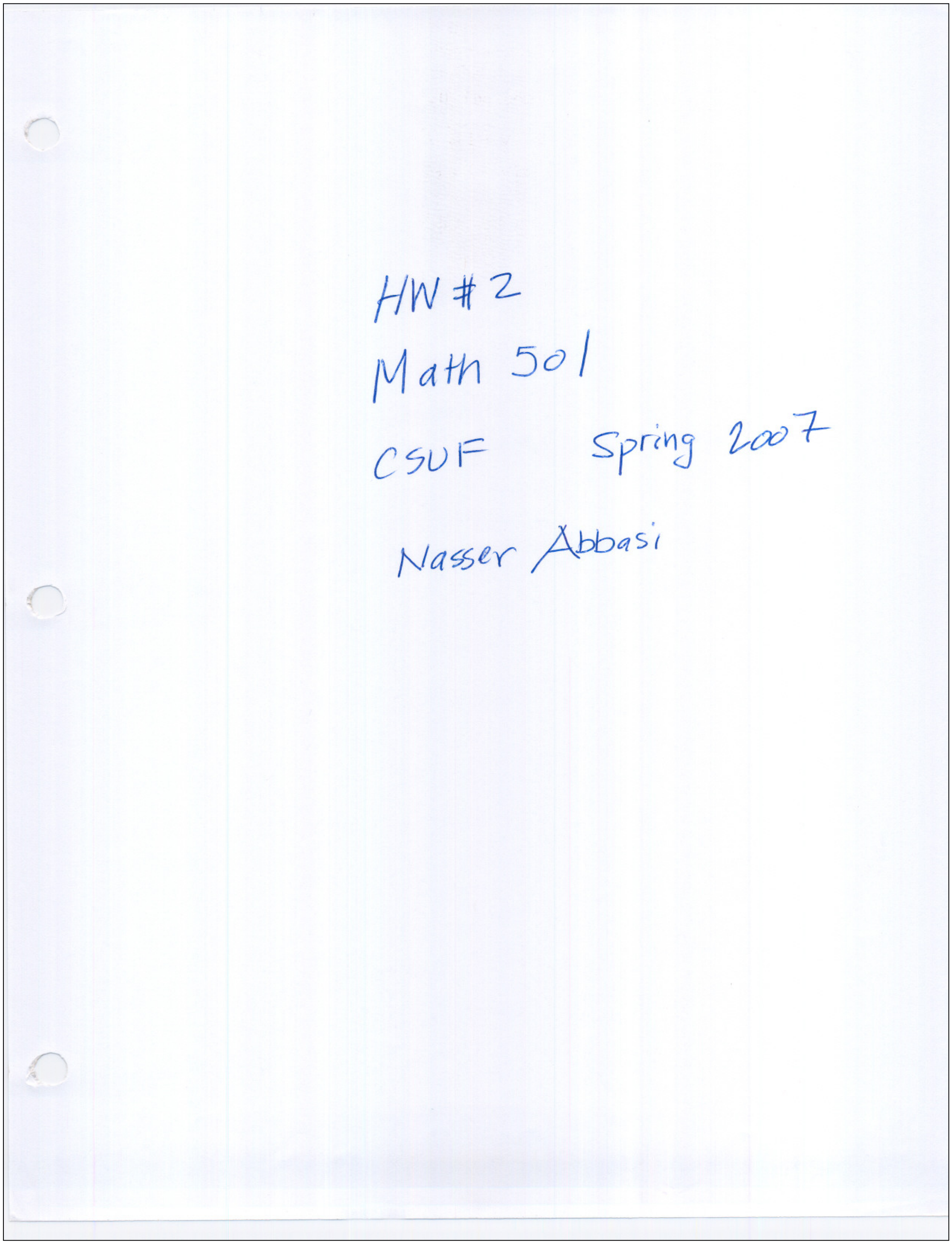
$$
\begin{aligned}
E^2 x_{n-2} - E x_{n-2} - E^0 x_{n-2} &= 0 \\
\left(E^2 - E - 1\right) x_{n-2} &= 0
\end{aligned}
$$

Hence the roots of the characteristic polynomial $p(E)x = 0$ are $\lambda^2 - \lambda - 1 = 0 \quad$ or $\lambda = \frac{-b \pm \sqrt{b^2-4ac}}{2a}$, hence $\lambda = \frac{1 \pm \sqrt{1+4}}{2} = \frac{1 \pm \sqrt{5}}{2}$

Hence $|\lambda_1| = \left|\frac{1+\sqrt{5}}{2}\right| = 1.618$ and $|\lambda_2| = \left|\frac{1-\sqrt{5}}{2}\right| = 0.618\,03$

Since $\lambda_1 \geq 1$, then $\boxed{\text{NOT STABLE}}$ difference equation.

4.3 HW 2

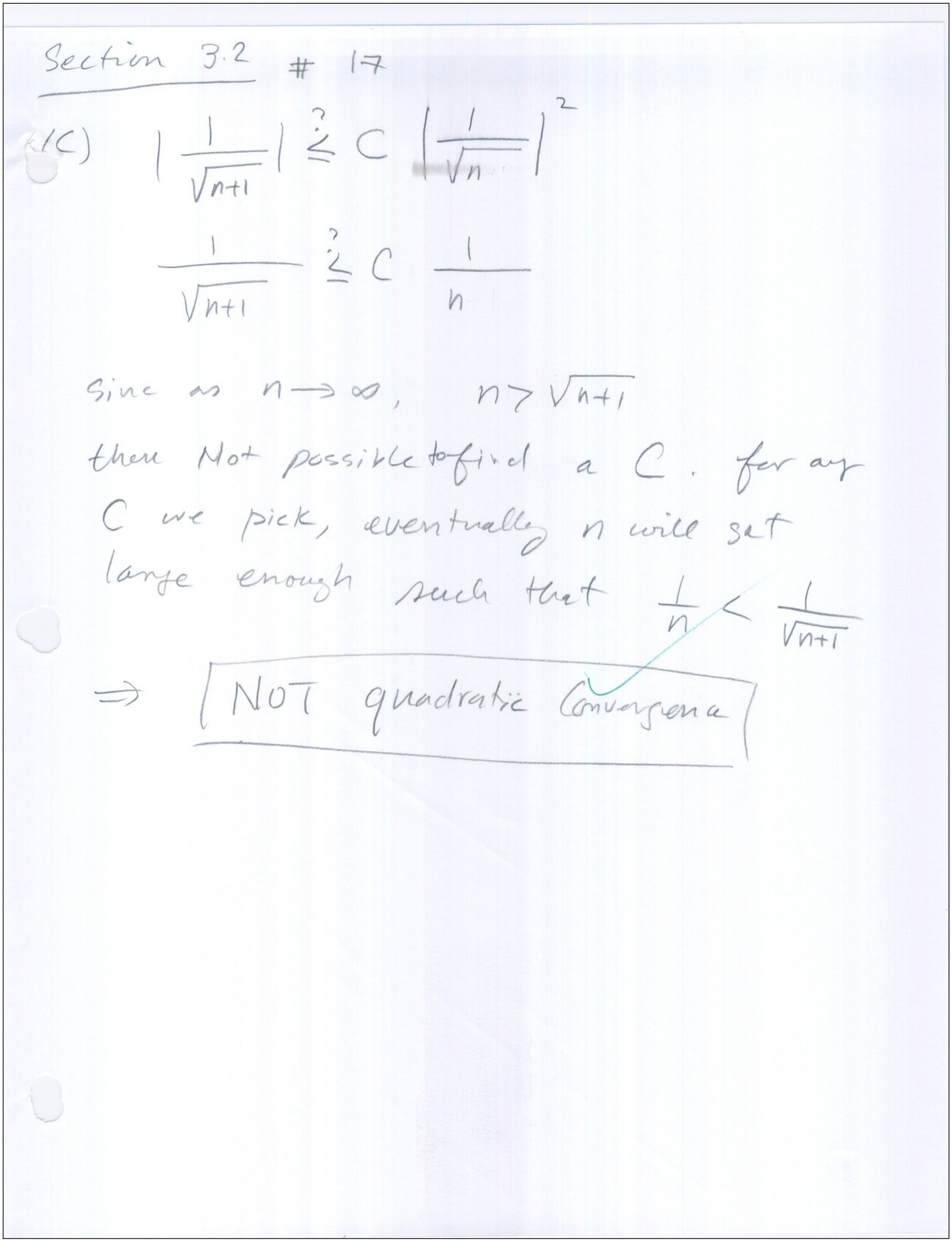


HW # 2

Math 501

CSUF Spring 2007

Nasser Abbasi



Section 3.2 # 17

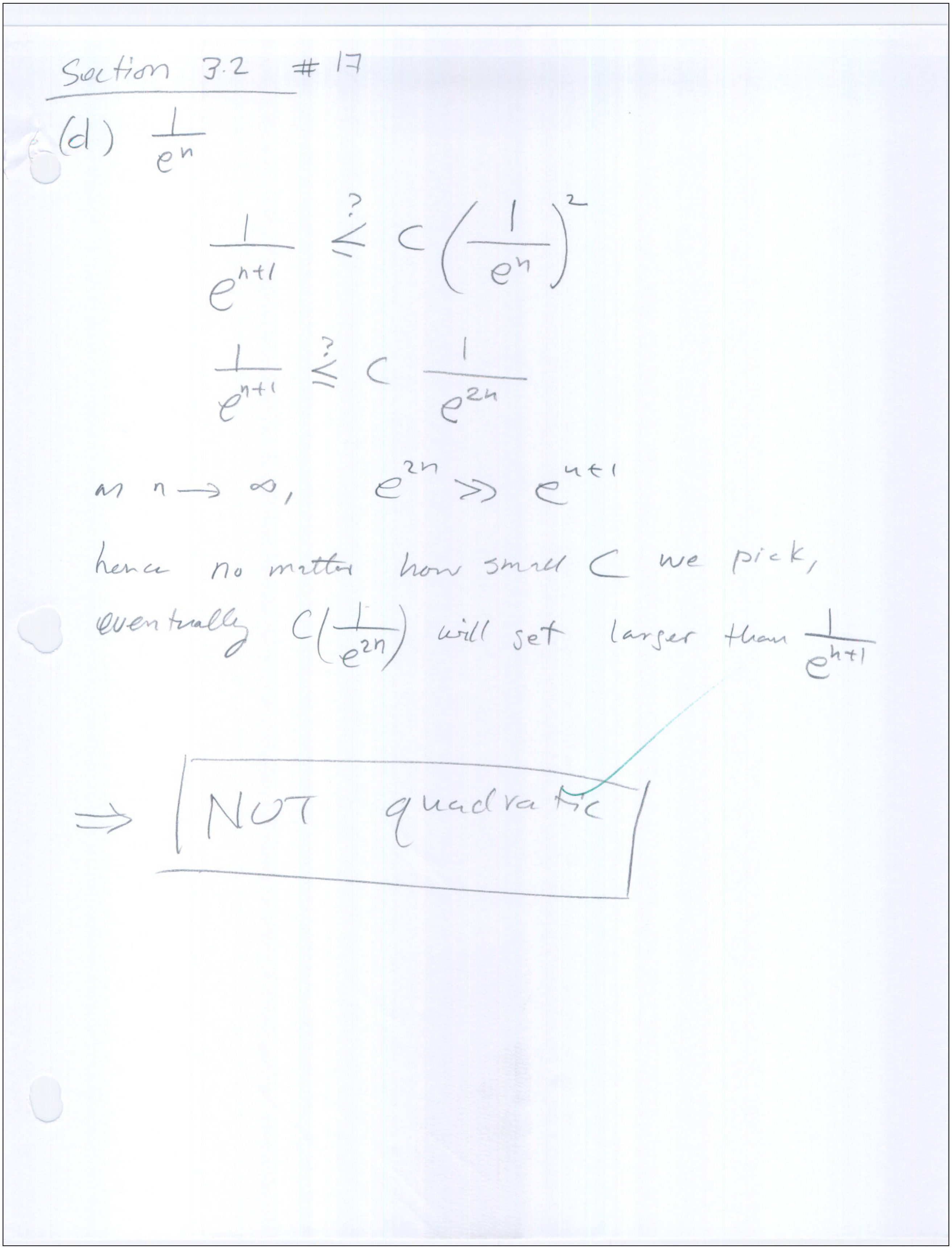
(C) $\left| \frac{1}{\sqrt{n+1}} \right| \overset{?}{\leq} C \left| \frac{1}{\sqrt{n}} \right|^2$

$$\frac{1}{\sqrt{n+1}} \overset{?}{\leq} C \frac{1}{n}$$

Since as $n \to \infty$, $n > \sqrt{n+1}$

then Not possible to find a C. for any C we pick, eventually n will get large enough such that $\frac{1}{n} < \frac{1}{\sqrt{n+1}}$

$\Rightarrow$ $\boxed{\text{NOT quadratic Convergence}}$



Section 3.2 # 17

(d) $\frac{1}{e^n}$

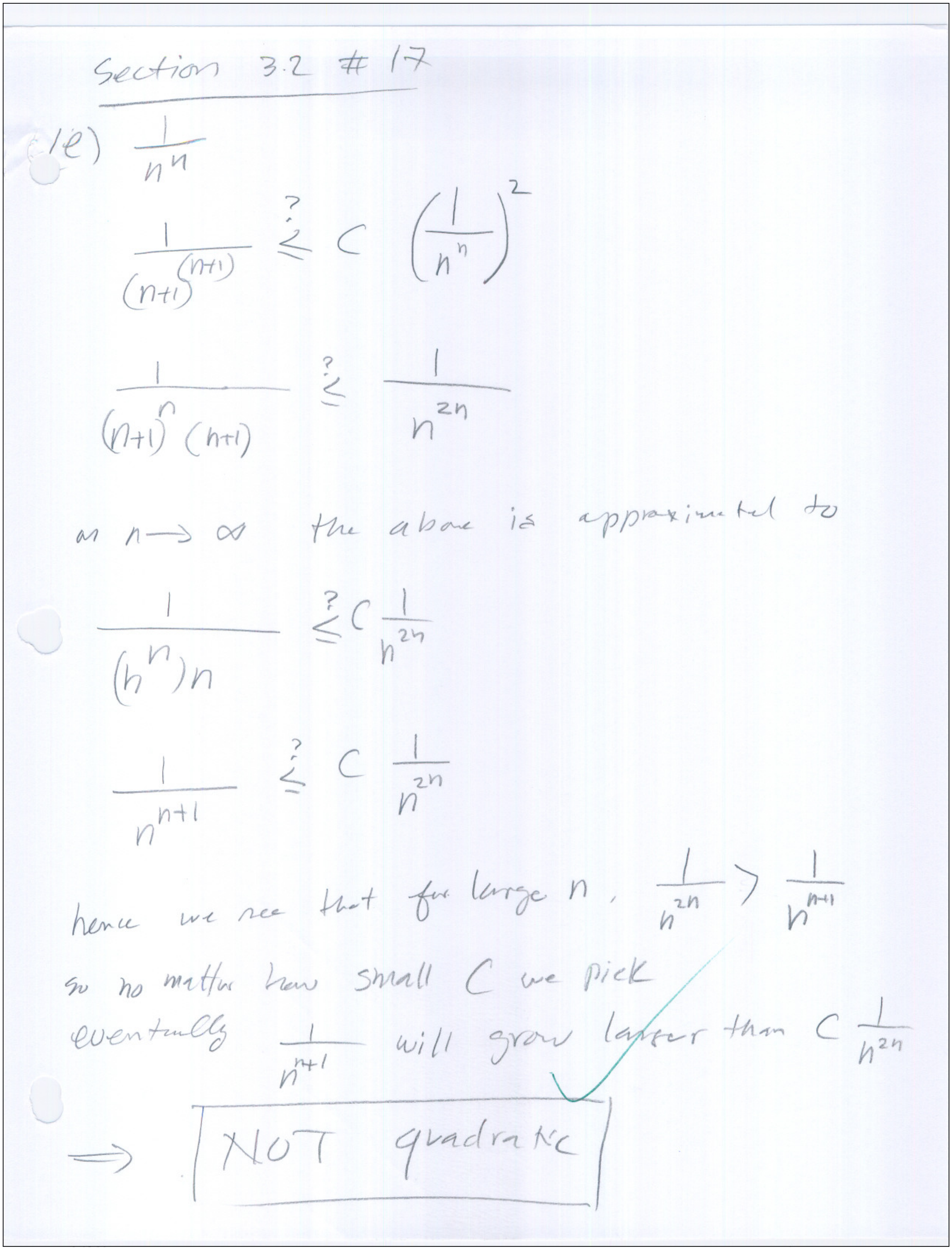
$$\frac{1}{e^{n+1}} \overset{?}{\leq} c\left(\frac{1}{e^n}\right)^2$$

$$\frac{1}{e^{n+1}} \overset{?}{\leq} c\,\frac{1}{e^{2n}}$$

as $n \longrightarrow \infty$, $e^{2n} \gg e^{n+1}$

hence no matter how small C we pick, eventually $C\left(\frac{1}{e^{2n}}\right)$ will get larger than $\frac{1}{e^{n+1}}$

$\Rightarrow$ $\boxed{\text{NOT quadratic}}$



Section 3.2 # 17

1e) $\frac{1}{n^n}$

$$\frac{1}{(n+1)^{(n+1)}} \overset{?}{\leq} C \left(\frac{1}{n^n}\right)^2$$

$$\frac{1}{(n+1)^n (n+1)} \overset{?}{\leq} \frac{1}{n^{2n}}$$

as $n \to \infty$ the above is approximated to

$$\frac{1}{(n^n)n} \overset{?}{\leq} C \frac{1}{n^{2n}}$$

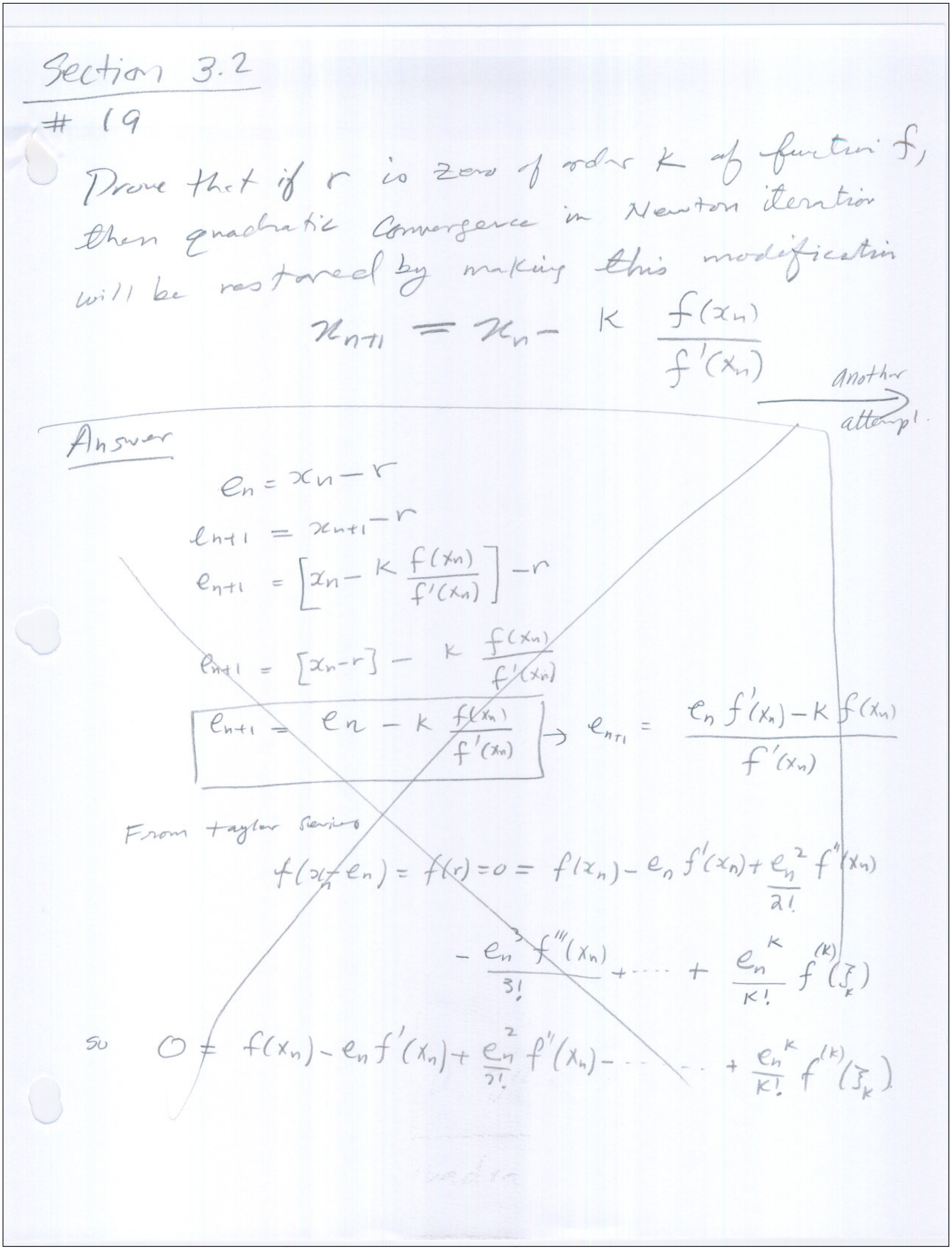
$$\frac{1}{n^{n+1}} \overset{?}{\leq} C \frac{1}{n^{2n}}$$

hence we see that for large n, $\frac{1}{n^{2n}} > \frac{1}{n^{n+1}}$

so no matter how small C we pick

eventually $\frac{1}{n^{n+1}}$ will grow larger than $C \frac{1}{n^{2n}}$

$\implies$ $\boxed{\text{NOT quadratic}}$



Section 3.2

# 19

Prove that if r is zero of order k of function f, then quadratic convergence in Newton iteration will be restored by making this modification

$$x_{n+1} = x_n - k\frac{f(x_n)}{f'(x_n)}$$

$\longrightarrow$ another attempt.

Answer

$$e_n = x_n - r$$

$$e_{n+1} = x_{n+1} - r$$

~~$e_{n+1} = \left[x_n - k\frac{f(x_n)}{f'(x_n)}\right] - r$~~

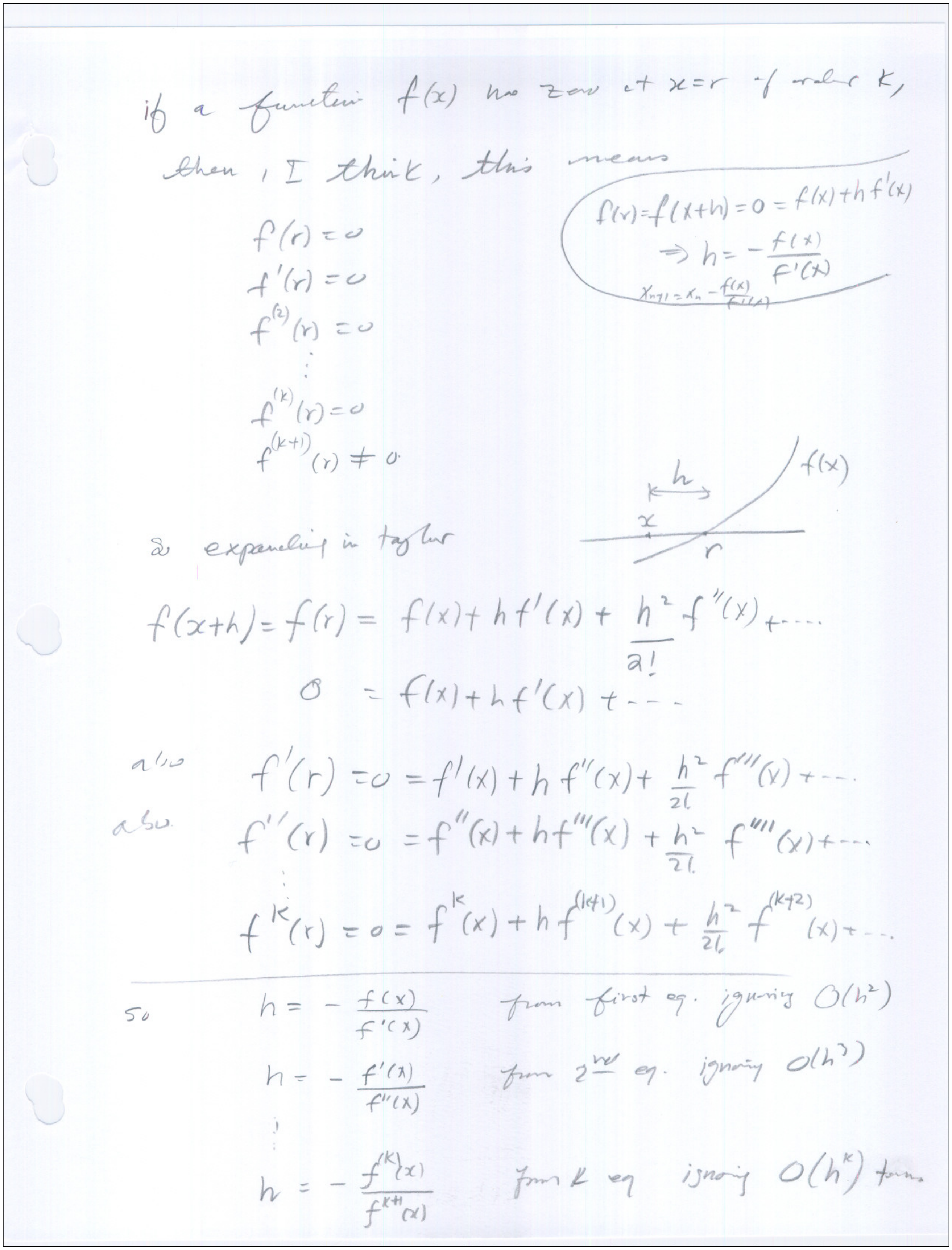
~~$e_{n+1} = [x_n - r] - k\frac{f(x_n)}{f'(x_n)}$~~

~~$\boxed{e_{n+1} = e_n - k\frac{f(x_n)}{f'(x_n)}} \rightarrow e_{n+1} = \frac{e_n f'(x_n) - k f(x_n)}{f'(x_n)}$~~

~~From taylor series~~

~~$f(x_n - e_n) = f(r) = 0 = f(x_n) - e_n f'(x_n) + \frac{e_n^2}{2!} f''(x_n) - \frac{e_n^3}{3!} f'''(x_n) + \cdots + \frac{e_n^k}{k!} f^{(k)}(\xi_k)$~~

~~so $0 = f(x_n) - e_n f'(x_n) + \frac{e_n^2}{2!} f''(x_n) - \cdots + \frac{e_n^k}{k!} f^{(k)}(\xi_k)$~~



if a function $f(x)$ has zero at $x=r$ of order k,

then, I think, this means

$$f(r)=0$$
$$f'(r)=0$$
$$f^{(2)}(r)=0$$
$$\vdots$$
$$f^{(k)}(r)=0$$
$$f^{(k+1)}(r)\neq 0$$

$$f(r)=f(x+h)=0=f(x)+hf'(x)$$
$$\Rightarrow h=-\frac{f(x)}{f'(x)}$$
$$x_{n+1}=x_n-\frac{f(x)}{f'(x)}$$

so expanding in taylor

$$f(x+h)=f(r)=f(x)+hf'(x)+\frac{h^2}{2!}f''(x)+\cdots$$
$$0=f(x)+hf'(x)+\cdots$$

also $$f'(r)=0=f'(x)+hf''(x)+\frac{h^2}{2!}f'''(x)+\cdots$$

also $$f''(r)=0=f''(x)+hf'''(x)+\frac{h^2}{2!}f''''(x)+\cdots$$

$$\vdots$$

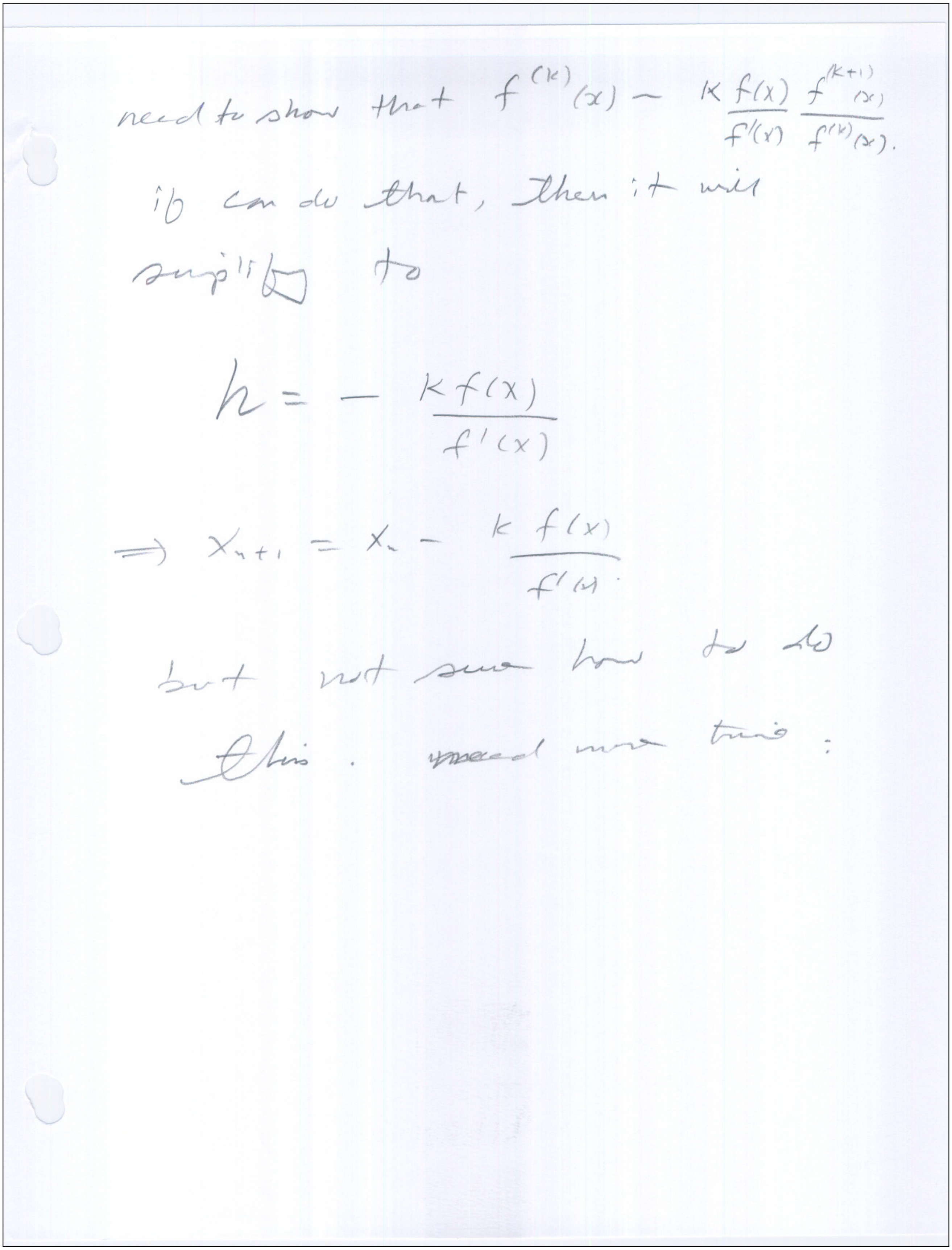
$$f^{k}(r)=0=f^{k}(x)+hf^{(k+1)}(x)+\frac{h^2}{2!}f^{(k+2)}(x)+\cdots$$

so $h=-\frac{f(x)}{f'(x)}$ from first eq. ignoring $O(h^2)$

$h=-\frac{f'(x)}{f''(x)}$ from 2^{nd} eq. ignoring $O(h^2)$

$\vdots$

$h=-\frac{f^{(k)}(x)}{f^{(k+1)}(x)}$ from k eq ignoring $O(h^k)$ terms



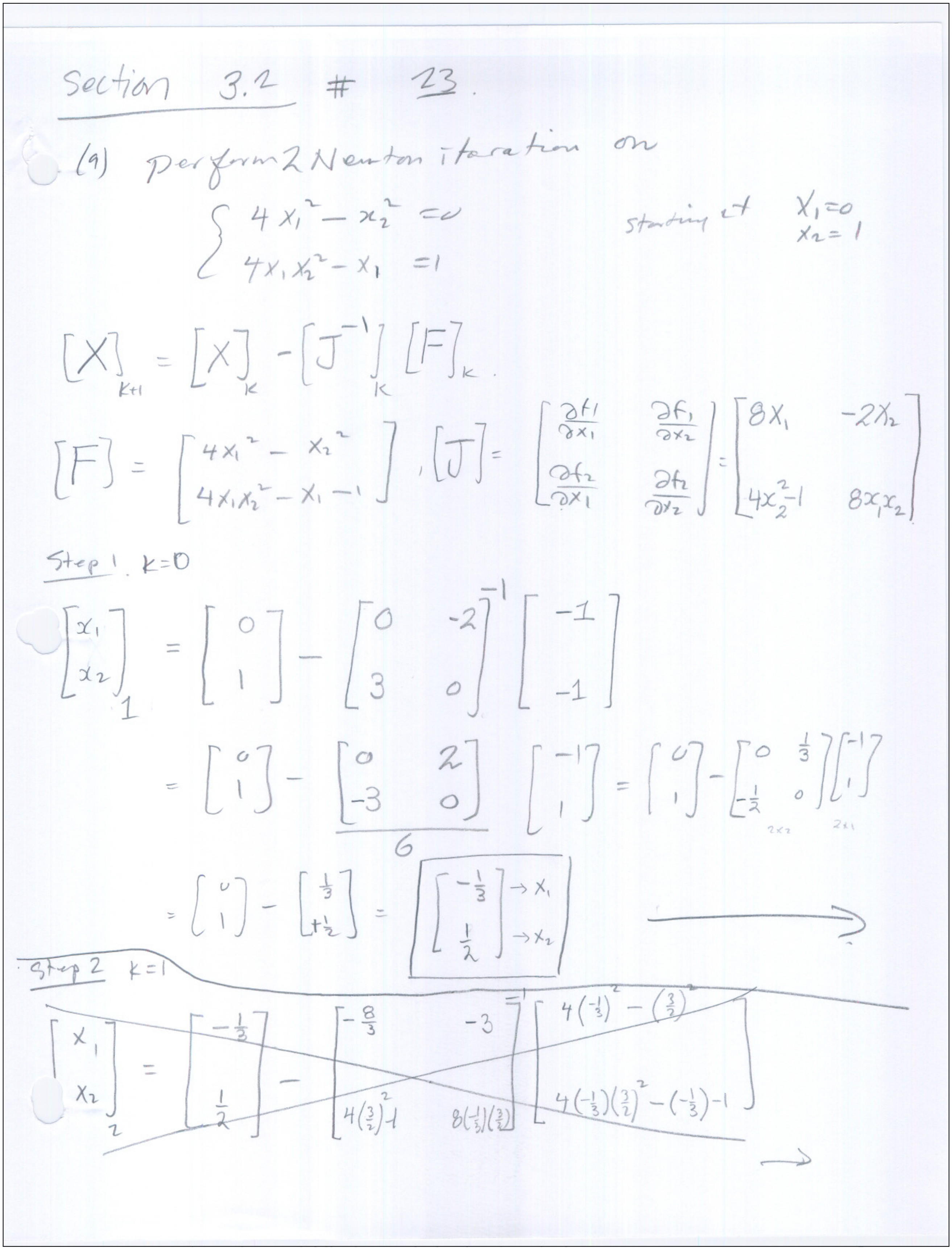
need to show that $f^{(k)}(x) \sim k\frac{f(x)}{f'(x)}\frac{f^{(k+1)}(x)}{f^{(k)}(x)}$.

if can do that, then it will simplify to

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

$$\Rightarrow x_{n+1} = x_n - \frac{k\,f(x)}{f'(x)}$$

but not sure how to do this. need more time:



Section 3.2 # 23.

(a) perform 2 Newton iteration on

$$\begin{cases} 4x_1^2 - x_2^2 = 0 \\ 4x_1x_2^2 - x_1 = 1 \end{cases}$$

starting at $x_1=0$, $x_2=1$

$$[X]_{k+1} = [X]_k - [J^{-1}]_k [F]_k .$$

$$[F] = \begin{bmatrix} 4x_1^2 - x_2^2 \\ 4x_1x_2^2 - x_1 - 1 \end{bmatrix}, \quad [J] = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix} = \begin{bmatrix} 8x_1 & -2x_2 \\ 4x_2^2-1 & 8x_1x_2 \end{bmatrix}$$

Step 1. k=0

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} 0 & -2 \\ 3 & 0 \end{bmatrix}^{-1} \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

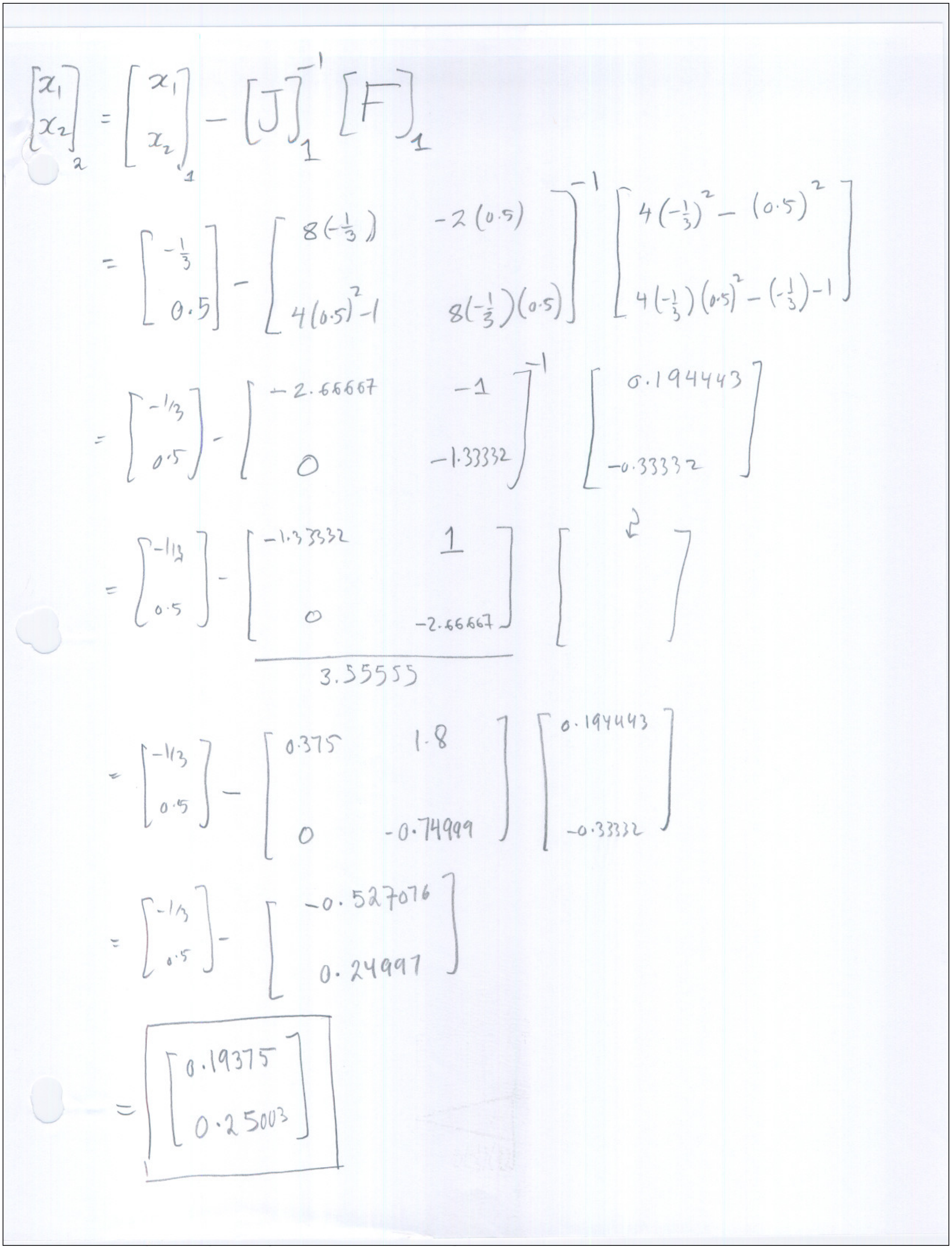
$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{\begin{bmatrix} 0 & 2 \\ -3 & 0 \end{bmatrix}}{6} \begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \underbrace{\begin{bmatrix} 0 & \frac{1}{3} \\ -\frac{1}{2} & 0 \end{bmatrix}}_{2\times 2} \underbrace{\begin{bmatrix} -1 \\ 1 \end{bmatrix}}_{2\times 1}$$

$$= \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \begin{bmatrix} \frac{1}{3} \\ +\frac{1}{2} \end{bmatrix} = \boxed{\begin{bmatrix} -\frac{1}{3} \\ \frac{1}{2} \end{bmatrix} \begin{matrix} \to x_1 \\ \to x_2 \end{matrix}} \longrightarrow$$

Step 2 k=1

~~$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}_2 = \begin{bmatrix} -\frac{1}{3} \\ \frac{1}{2} \end{bmatrix} - \begin{bmatrix} -\frac{8}{3} & -3 \\ 4(\frac{3}{2})^2-1 & 8(-\frac{1}{3})(\frac{3}{2}) \end{bmatrix}^{-1} \begin{bmatrix} 4(-\frac{1}{3})^2 - (\frac{3}{2})^2 \\ 4(-\frac{1}{3})(\frac{3}{2})^2 - (-\frac{1}{3}) - 1 \end{bmatrix}$~~

$\longrightarrow$



$$\begin{bmatrix}x_1\\x_2\end{bmatrix}_2 = \begin{bmatrix}x_1\\x_2\end{bmatrix}_1 - [J]_1^{-1}\,[F]_1$$

$$= \begin{bmatrix}-\frac{1}{3}\\0.5\end{bmatrix} - \begin{bmatrix}8(-\frac{1}{3}) & -2(0.5)\\4(0.5)^2-1 & 8(-\frac{1}{3})(0.5)\end{bmatrix}^{-1}\begin{bmatrix}4(-\frac{1}{3})^2-(0.5)^2\\4(-\frac{1}{3})(0.5)^2-(-\frac{1}{3})-1\end{bmatrix}$$

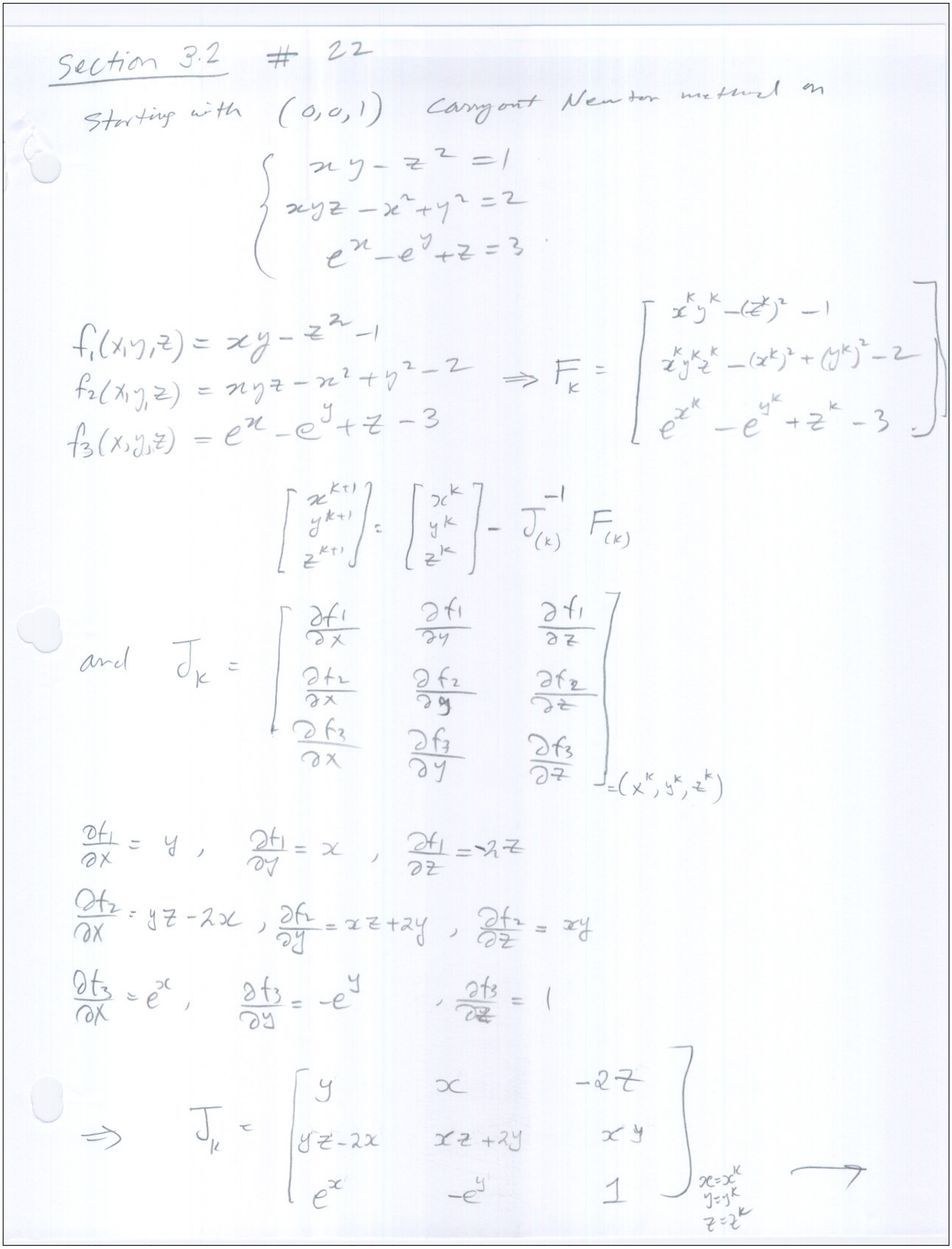
$$= \begin{bmatrix}-1/3\\0.5\end{bmatrix} - \begin{bmatrix}-2.66667 & -1\\0 & -1.33332\end{bmatrix}^{-1}\begin{bmatrix}0.194443\\-0.33332\end{bmatrix}$$

$$= \begin{bmatrix}-1/3\\0.5\end{bmatrix} - \frac{\begin{bmatrix}-1.33332 & 1\\0 & -2.66667\end{bmatrix}}{3.55555}\begin{bmatrix}\hookleftarrow\\ \\ \end{bmatrix}$$

$$= \begin{bmatrix}-1/3\\0.5\end{bmatrix} - \begin{bmatrix}0.375 & 1.8\\0 & -0.74999\end{bmatrix}\begin{bmatrix}0.194443\\-0.33332\end{bmatrix}$$

$$= \begin{bmatrix}-1/3\\0.5\end{bmatrix} - \begin{bmatrix}-0.527076\\0.24997\end{bmatrix}$$

$$= \boxed{\begin{bmatrix}0.19375\\0.25003\end{bmatrix}}$$



Section 3.2 # 22

Starting with $(0,0,1)$ carry out Newton method on

$$\begin{cases} xy - z^2 = 1 \\ xyz - x^2 + y^2 = 2 \\ e^x - e^y + z = 3 \end{cases}$$

$$f_1(x,y,z) = xy - z^2 - 1$$
$$f_2(x,y,z) = xyz - x^2 + y^2 - 2 \quad \Rightarrow F_k = \begin{bmatrix} x^k y^k - (z^k)^2 - 1 \\ x^k y^k z^k - (x^k)^2 + (y^k)^2 - 2 \\ e^{x^k} - e^{y^k} + z^k - 3 \end{bmatrix}$$
$$f_3(x,y,z) = e^x - e^y + z - 3$$

$$\begin{bmatrix} x^{k+1} \\ y^{k+1} \\ z^{k+1} \end{bmatrix} = \begin{bmatrix} x^k \\ y^k \\ z^k \end{bmatrix} - J_{(k)}^{-1} F_{(k)}$$

and

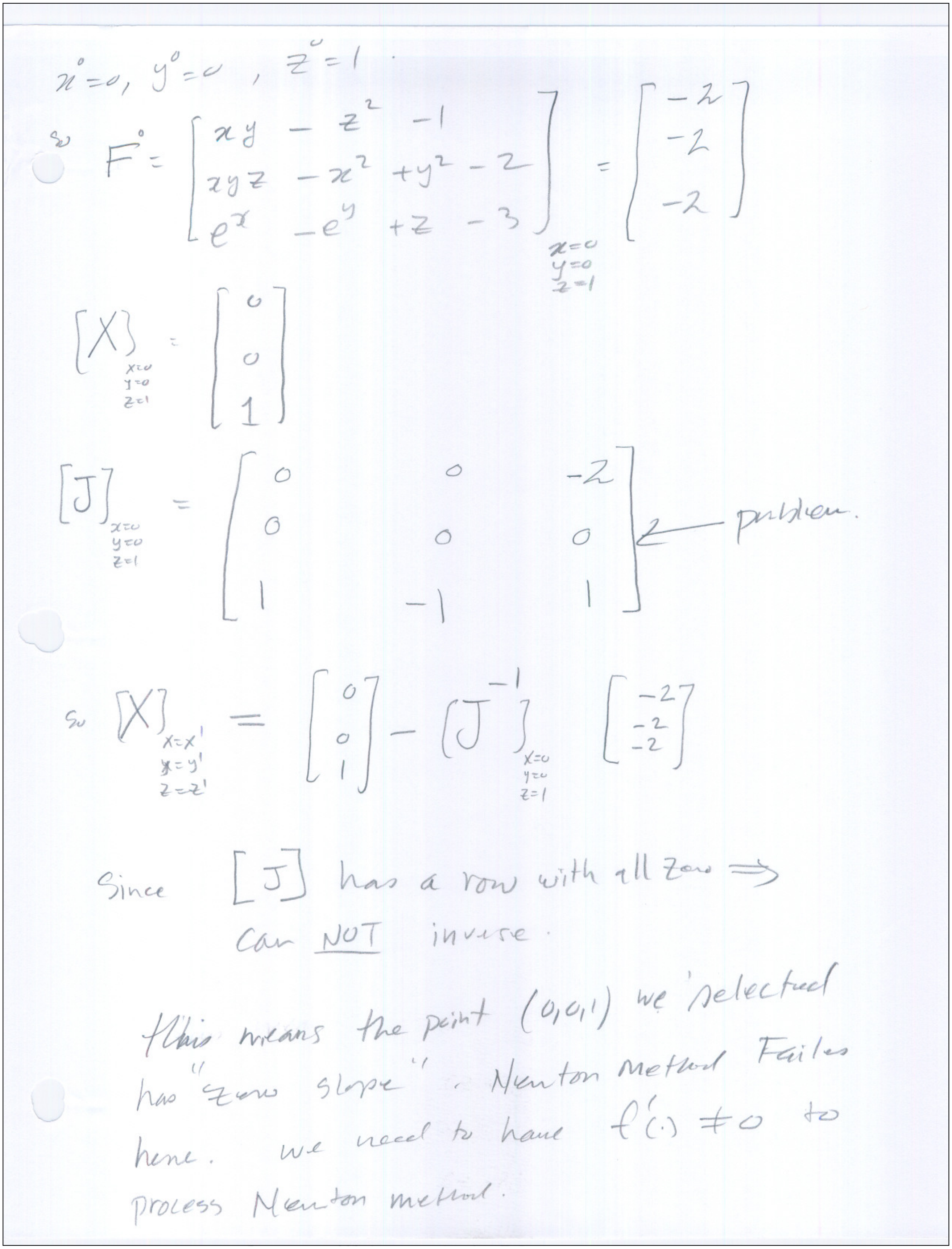
$$J_k = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} & \frac{\partial f_1}{\partial z} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} & \frac{\partial f_2}{\partial z} \\ \frac{\partial f_3}{\partial x} & \frac{\partial f_3}{\partial y} & \frac{\partial f_3}{\partial z} \end{bmatrix}_{=(x^k, y^k, z^k)}$$

$$\frac{\partial f_1}{\partial x} = y, \quad \frac{\partial f_1}{\partial y} = x, \quad \frac{\partial f_1}{\partial z} = -2z$$

$$\frac{\partial f_2}{\partial x} = yz - 2x, \quad \frac{\partial f_2}{\partial y} = xz + 2y, \quad \frac{\partial f_2}{\partial z} = xy$$

$$\frac{\partial f_3}{\partial x} = e^x, \quad \frac{\partial f_3}{\partial y} = -e^y, \quad \frac{\partial f_3}{\partial z} = 1$$

$$\Rightarrow \quad J_k = \begin{bmatrix} y & x & -2z \\ yz - 2x & xz + 2y & xy \\ e^x & -e^y & 1 \end{bmatrix}_{\substack{x = x^k \\ y = y^k \\ z = z^k}} \longrightarrow$$



$x^0=0,\ y^0=0\ ,\ z^0=1$.

so $$F^0 = \begin{bmatrix} xy & -z^2 & -1 \\ xyz & -x^2 & +y^2-2 \\ e^x & -e^y & +z-3 \end{bmatrix}_{\substack{x=0\\y=0\\z=1}} = \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix}$$

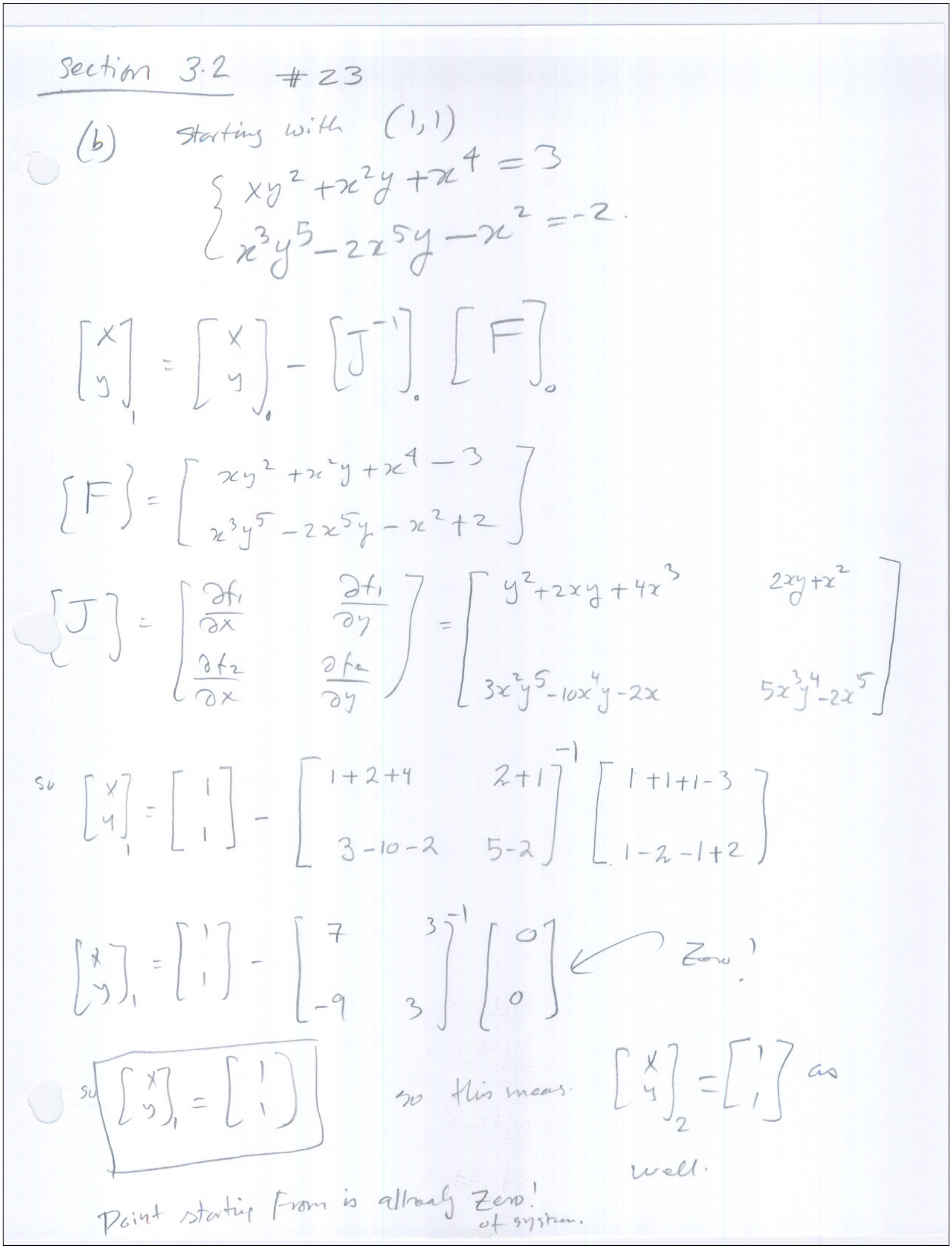
$$[X]_{\substack{x=0\\y=0\\z=1}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

$$[J]_{\substack{x=0\\y=0\\z=1}} = \begin{bmatrix} 0 & 0 & -2 \\ 0 & 0 & 0 \\ 1 & -1 & 1 \end{bmatrix} \leftarrow \text{problem.}$$

so $$[X]_{\substack{x=x'\\y=y'\\z=z'}} = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} - \left([J]^{-1}\right)_{\substack{x=0\\y=0\\z=1}} \begin{bmatrix} -2 \\ -2 \\ -2 \end{bmatrix}$$

Since $[J]$ has a row with all zero $\Longrightarrow$ can NOT inverse.

this means the point $(0,0,1)$ we selected has "zero slope". Newton Method Failes here. we need to have $f'(\cdot) \neq 0$ to process Newton method.



Section 3.2 #23

(b) Starting with (1,1)

$$\begin{cases} xy^2+x^2y+x^4=3 \\ x^3y^5-2x^5y-x^2=-2. \end{cases}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}_1 = \begin{bmatrix} x \\ y \end{bmatrix}_0 - \begin{bmatrix} J^{-1} \end{bmatrix}_0 \begin{bmatrix} F \end{bmatrix}_0$$

$$\{F\} = \begin{bmatrix} xy^2+x^2y+x^4-3 \\ x^3y^5-2x^5y-x^2+2 \end{bmatrix}$$

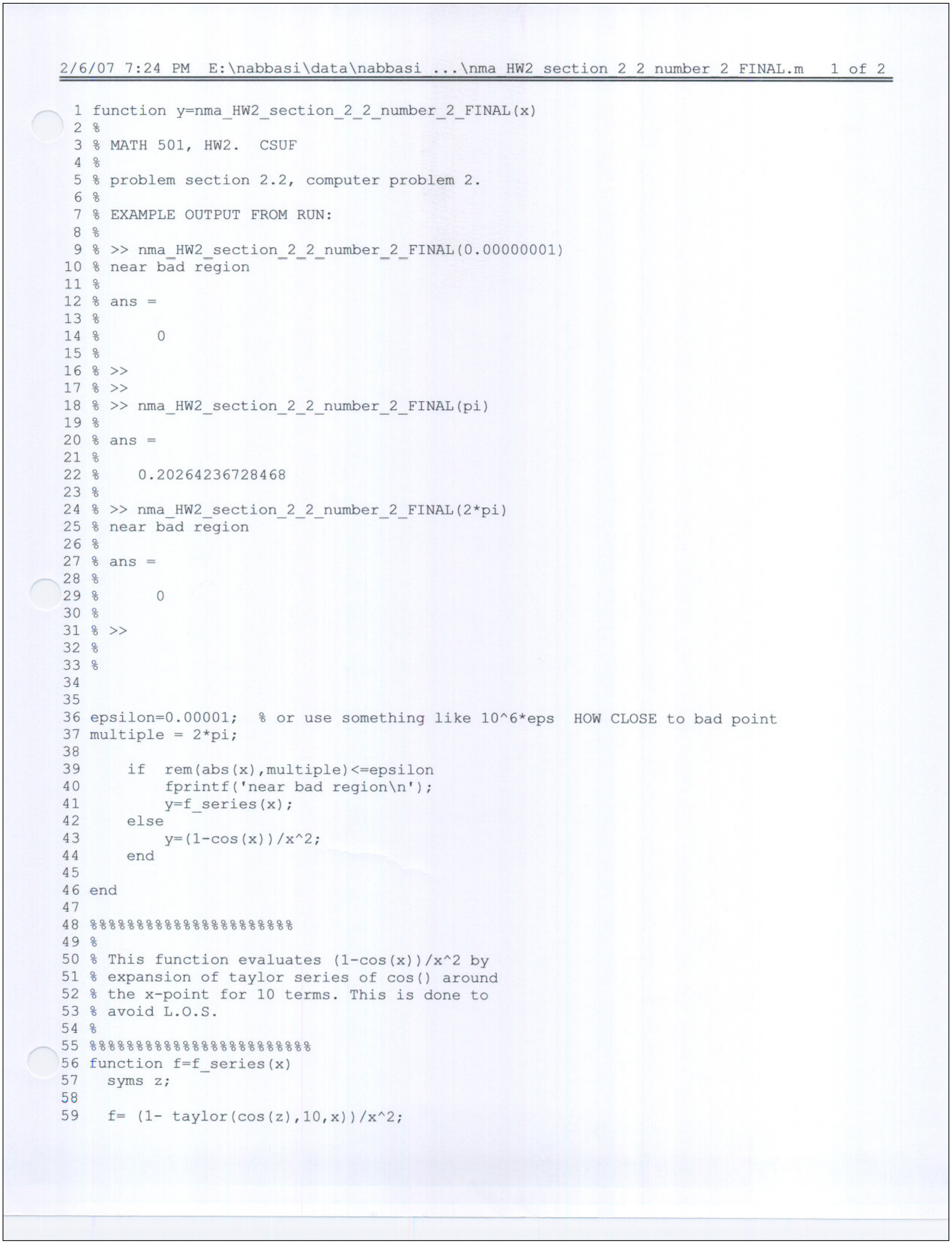
$$[J] = \begin{bmatrix} \frac{\partial f_1}{\partial x} & \frac{\partial f_1}{\partial y} \\ \frac{\partial f_2}{\partial x} & \frac{\partial f_2}{\partial y} \end{bmatrix} = \begin{bmatrix} y^2+2xy+4x^3 & 2xy+x^2 \\ 3x^2y^5-10x^4y-2x & 5x^3y^4-2x^5 \end{bmatrix}$$

so $$\begin{bmatrix} x \\ y \end{bmatrix}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 1+2+4 & 2+1 \\ 3-10-2 & 5-2 \end{bmatrix}^{-1} \begin{bmatrix} 1+1+1-3 \\ 1-2-1+2 \end{bmatrix}$$

$$\begin{bmatrix} x \\ y \end{bmatrix}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \begin{bmatrix} 7 & 3 \\ -9 & 3 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \end{bmatrix} \longleftarrow \text{Zero!}$$

so $\boxed{\begin{bmatrix} x \\ y \end{bmatrix}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}}$ so this means $\begin{bmatrix} x \\ y \end{bmatrix}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ as well.

Point starting from is already Zero! of system.

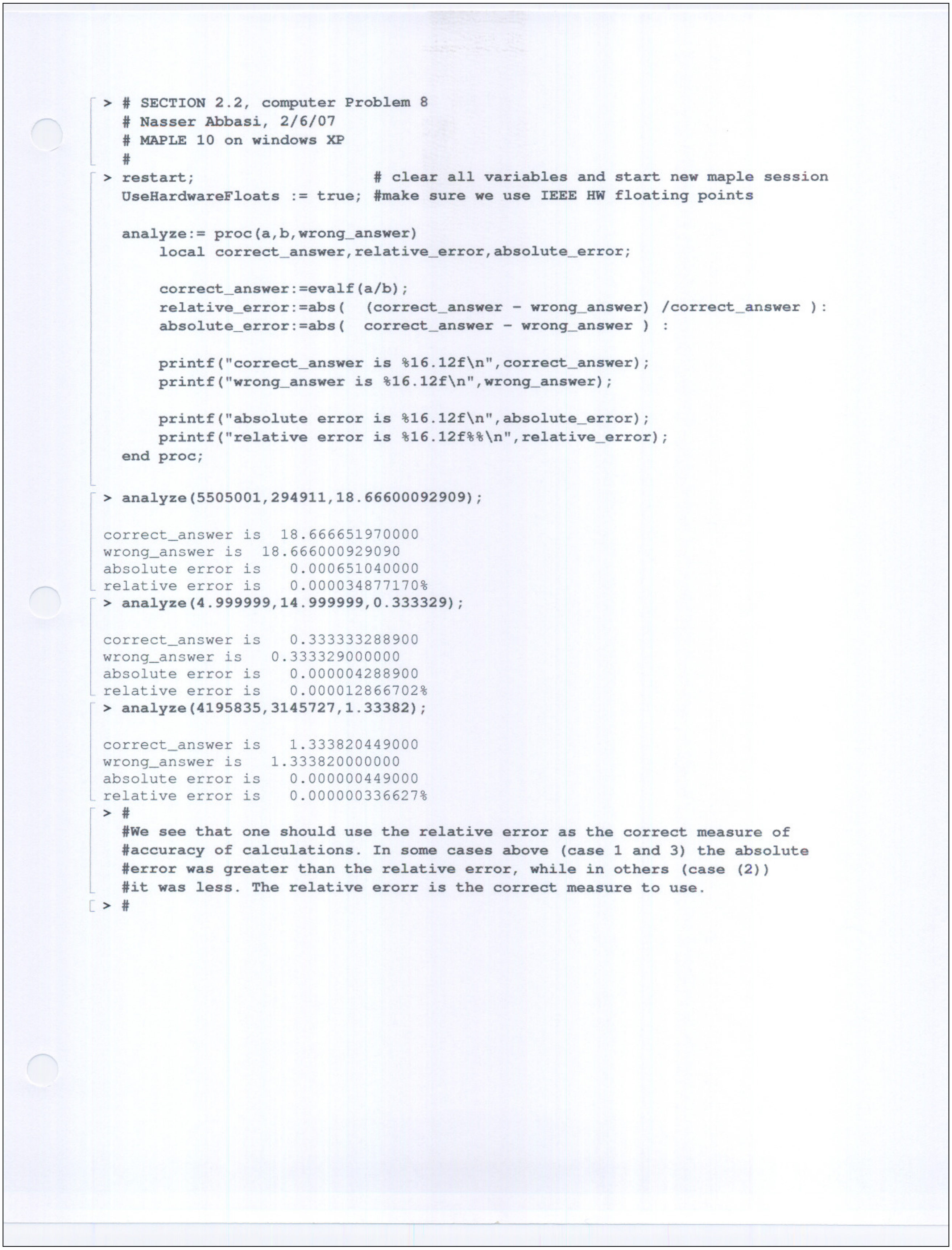


2/6/07 7:24 PM E:\nabbasi\data\nabbasi ...\nma HW2 section 2 2 number 2 FINAL.m 1 of 2

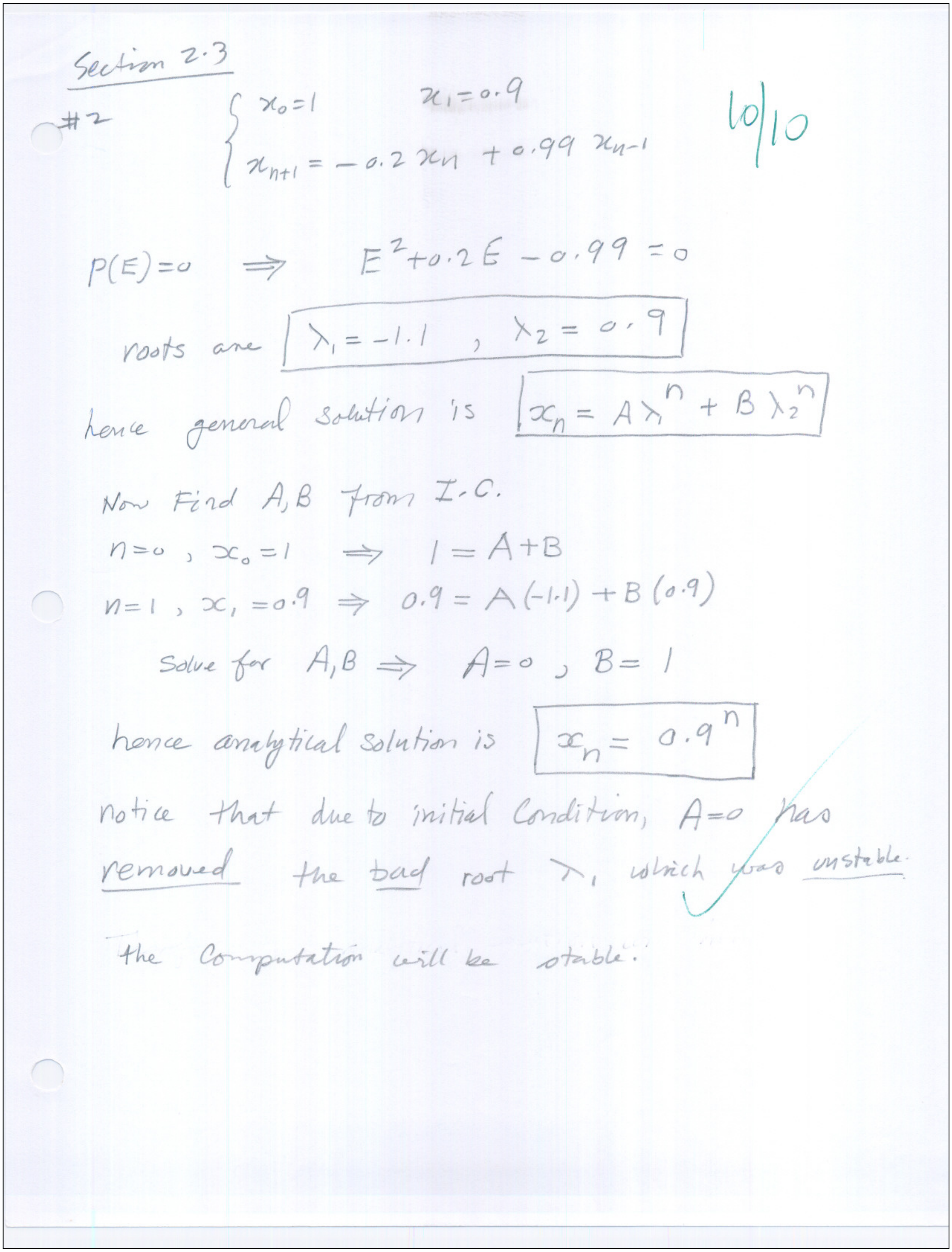
```
function y=nma_HW2_section_2_2_number_2_FINAL(x)
%
% MATH 501, HW2.  CSUF
%
% problem section 2.2, computer problem 2.
%
% EXAMPLE OUTPUT FROM RUN:
%
% >> nma_HW2_section_2_2_number_2_FINAL(0.00000001)
% near bad region
%
% ans =
%
%      0
%
% >>
% >>
% >> nma_HW2_section_2_2_number_2_FINAL(pi)
%
% ans =
%
%    0.20264236728468
%
% >> nma_HW2_section_2_2_number_2_FINAL(2*pi)
% near bad region
%
% ans =
%
%      0
%
% >>
%
%


epsilon=0.00001;  % or use something like 10^6*eps  HOW CLOSE to bad point
multiple = 2*pi;

    if  rem(abs(x),multiple)<=epsilon
        fprintf('near bad region\n');
        y=f_series(x);
    else
        y=(1-cos(x))/x^2;
    end

end

%%%%%%%%%%%%%%%%%%%%%%
%
% This function evaluates (1-cos(x))/x^2 by
% expansion of taylor series of cos() around
% the x-point for 10 terms. This is done to
% avoid L.O.S.
%
%%%%%%%%%%%%%%%%%%%%%%%%
function f=f_series(x)
  syms z;

  f= (1- taylor(cos(z),10,x))/x^2;
```

2/6/07 7:24 PM E:\nabbasi\data\nabbasi ...\nma HW2 section 2 2 number 2 FINAL.m 2 of 2

```
  f=double(subs(f,z,x));
end

```



```
> # SECTION 2.2, computer Problem 8
  # Nasser Abbasi, 2/6/07
  # MAPLE 10 on windows XP
  #
> restart;                   # clear all variables and start new maple session
  UseHardwareFloats := true; #make sure we use IEEE HW floating points

  analyze:= proc(a,b,wrong_answer)
      local correct_answer,relative_error,absolute_error;

      correct_answer:=evalf(a/b);
      relative_error:=abs(  (correct_answer - wrong_answer) /correct_answer ):
      absolute_error:=abs(  correct_answer - wrong_answer ) :

      printf("correct_answer is %16.12f\n",correct_answer);
      printf("wrong_answer is %16.12f\n",wrong_answer);

      printf("absolute error is %16.12f\n",absolute_error);
      printf("relative error is %16.12f%%\n",relative_error);
  end proc;

> analyze(5505001,294911,18.66600092909);

correct_answer is  18.666651970000
wrong_answer is  18.666000929090
absolute error is   0.000651040000
relative error is   0.000034877170%
> analyze(4.999999,14.999999,0.333329);

correct_answer is   0.333333288900
wrong_answer is   0.333329000000
absolute error is   0.000004288900
relative error is   0.000012866702%
> analyze(4195835,3145727,1.33382);

correct_answer is   1.333820449000
wrong_answer is   1.333820000000
absolute error is   0.000000449000
relative error is   0.000000336627%
> #
  #We see that one should use the relative error as the correct measure of
  #accuracy of calculations. In some cases above (case 1 and 3) the absolute
  #error was greater than the relative error, while in others (case (2))
  #it was less. The relative erorr is the correct measure to use.
> #
```



Section 2.3

10/10

#2 $\begin{cases} x_0 = 1 & x_1 = 0.9 \\ x_{n+1} = -0.2\, x_n + 0.99\, x_{n-1} \end{cases}$

$P(E) = 0 \Rightarrow E^2 + 0.2E - 0.99 = 0$

roots are $\boxed{\lambda_1 = -1.1 \quad , \quad \lambda_2 = 0.9}$

hence general solution is $\boxed{x_n = A\lambda_1^n + B\lambda_2^n}$

Now Find A, B from I.C.

$n=0 \ , \ x_0 = 1 \Rightarrow 1 = A + B$

$n=1 \ , \ x_1 = 0.9 \Rightarrow 0.9 = A(-1.1) + B(0.9)$

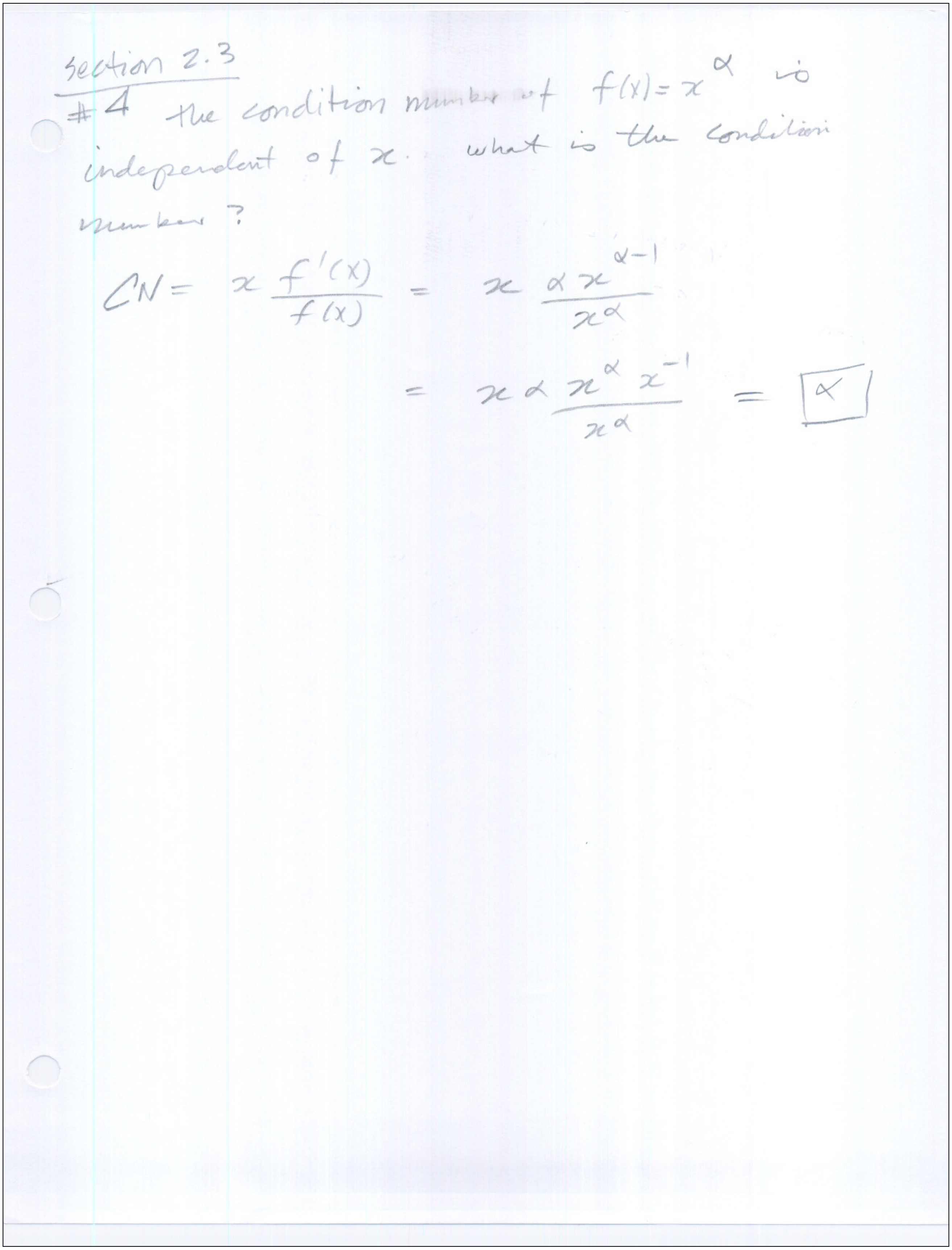
Solve for $A, B \Rightarrow A = 0 \ , \ B = 1$

hence analytical solution is $\boxed{x_n = 0.9^n}$

notice that due to initial Condition, $A=0$ has removed the bad root λ_1 which was unstable.

✓

the Computation will be stable.

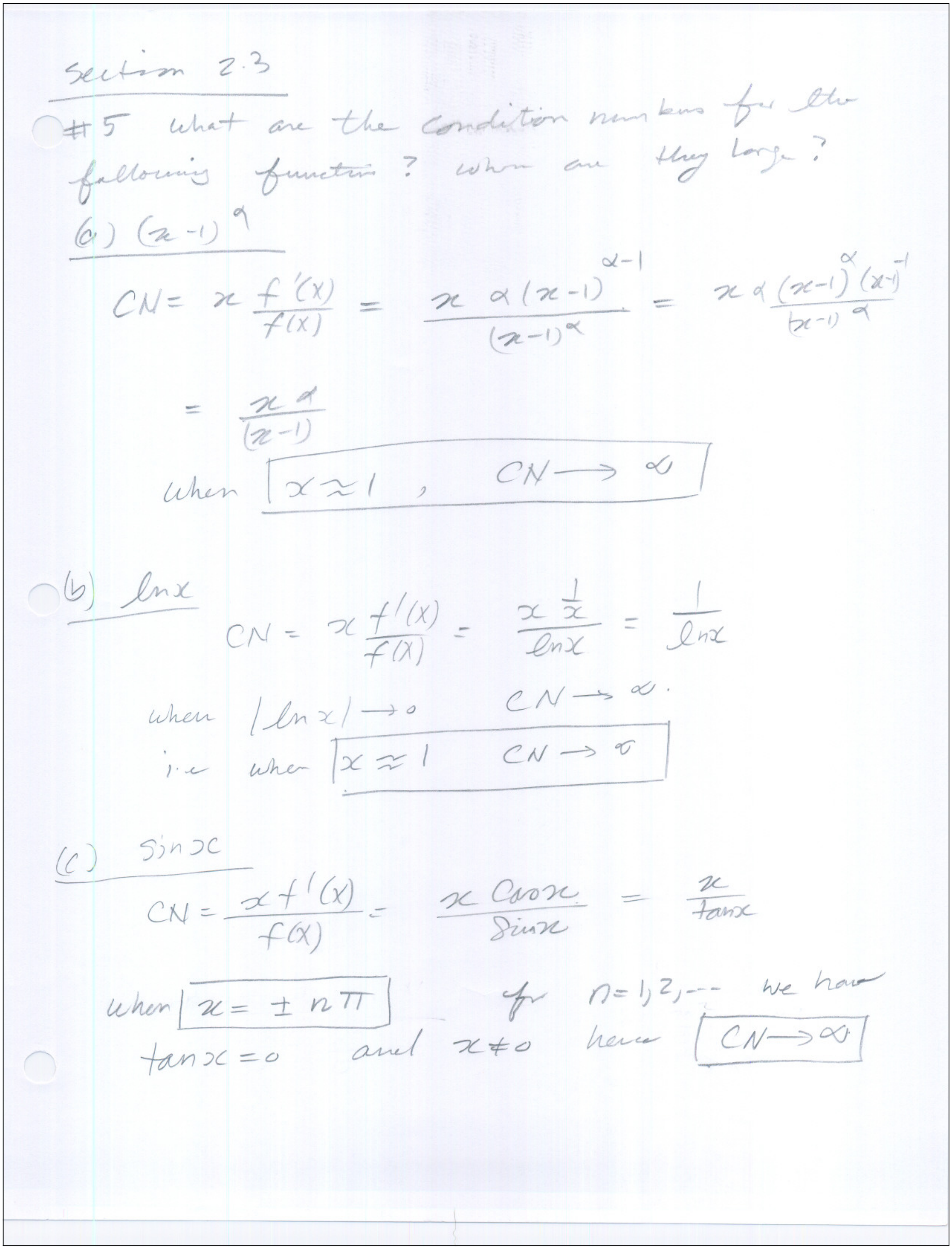


section 2.3

#4 the condition number of $f(x) = x^\alpha$ is independent of x. what is the condition number?

$$CN = x\frac{f'(x)}{f(x)} = x\frac{\alpha x^{\alpha-1}}{x^\alpha}$$

$$= x\alpha\frac{x^\alpha x^{-1}}{x^\alpha} = \boxed{\alpha}$$



section 2.3

#5 what are the condition numbers for the following functions? when are they large?

(a) $(x-1)^\alpha$

$$CN = x\frac{f'(x)}{f(x)} = \frac{x\,\alpha(x-1)^{\alpha-1}}{(x-1)^\alpha} = x\alpha\frac{(x-1)^\alpha(x-1)^{-1}}{(x-1)^\alpha}$$

$$= \frac{x\alpha}{(x-1)}$$

when $\boxed{x \approx 1 , \quad CN \to \infty}$

(b) $\ln x$

$$CN = x\frac{f'(x)}{f(x)} = \frac{x\frac{1}{x}}{\ln x} = \frac{1}{\ln x}$$

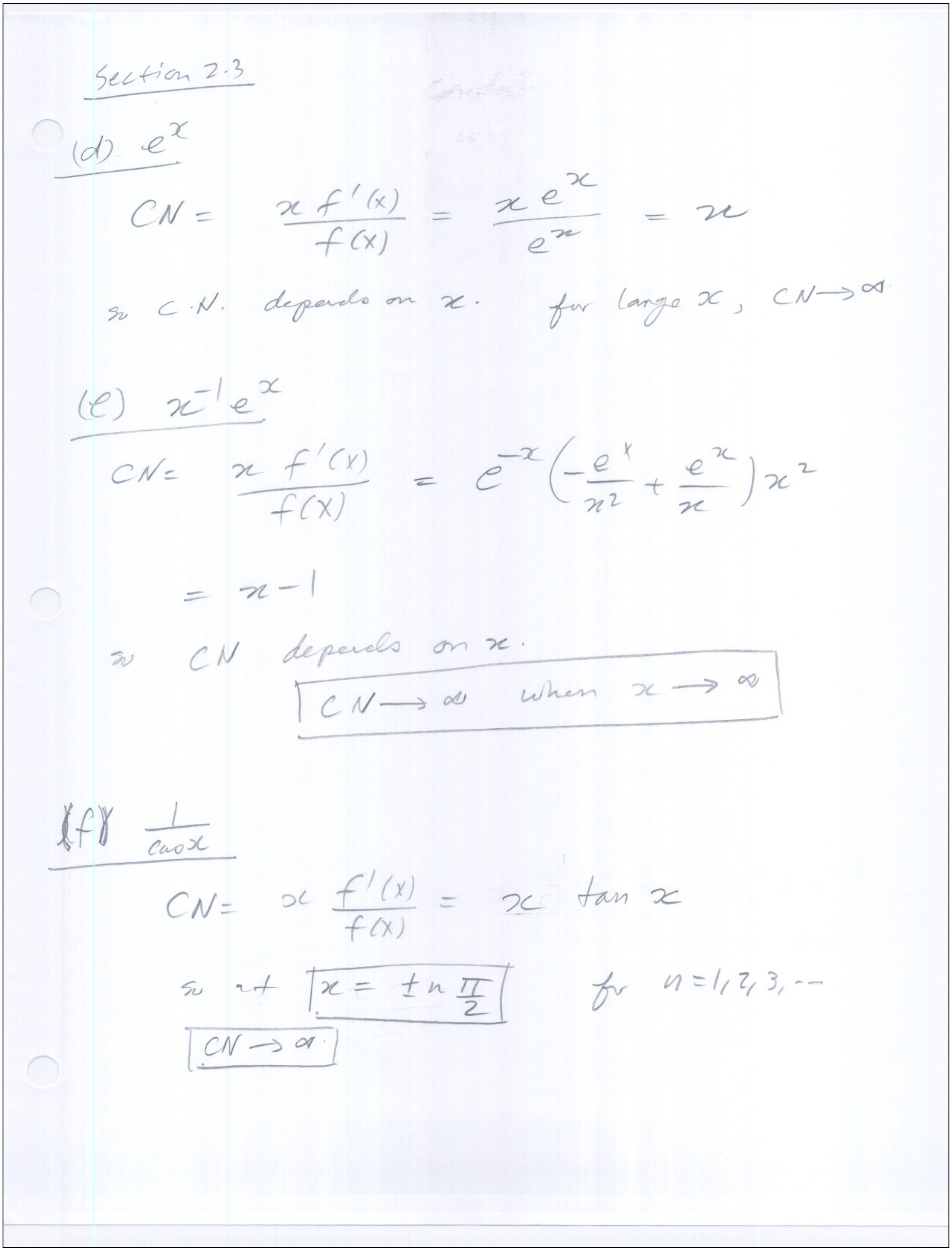
when $|\ln x| \to 0$ $CN \to \infty$.

i.e when $\boxed{x \approx 1 \quad CN \to \infty}$

(c) $\sin x$

$$CN = \frac{xf'(x)}{f(x)} = \frac{x\cos x}{\sin x} = \frac{x}{\tan x}$$

when $\boxed{x = \pm n\pi}$ for $n = 1, 2, \dots$ we have $\tan x = 0$ and $x \neq 0$ hence $\boxed{CN \to \infty}$



Section 2.3

(d) e^x

$$CN = \frac{x f'(x)}{f(x)} = \frac{x e^x}{e^x} = x$$

so C.N. depends on x. for large x, $CN \to \infty$.

(e) $x^{-1} e^x$

$$CN = \frac{x f'(x)}{f(x)} = e^{-x}\left(-\frac{e^x}{x^2} + \frac{e^x}{x}\right)x^2$$

$$= x - 1$$

so CN depends on x.

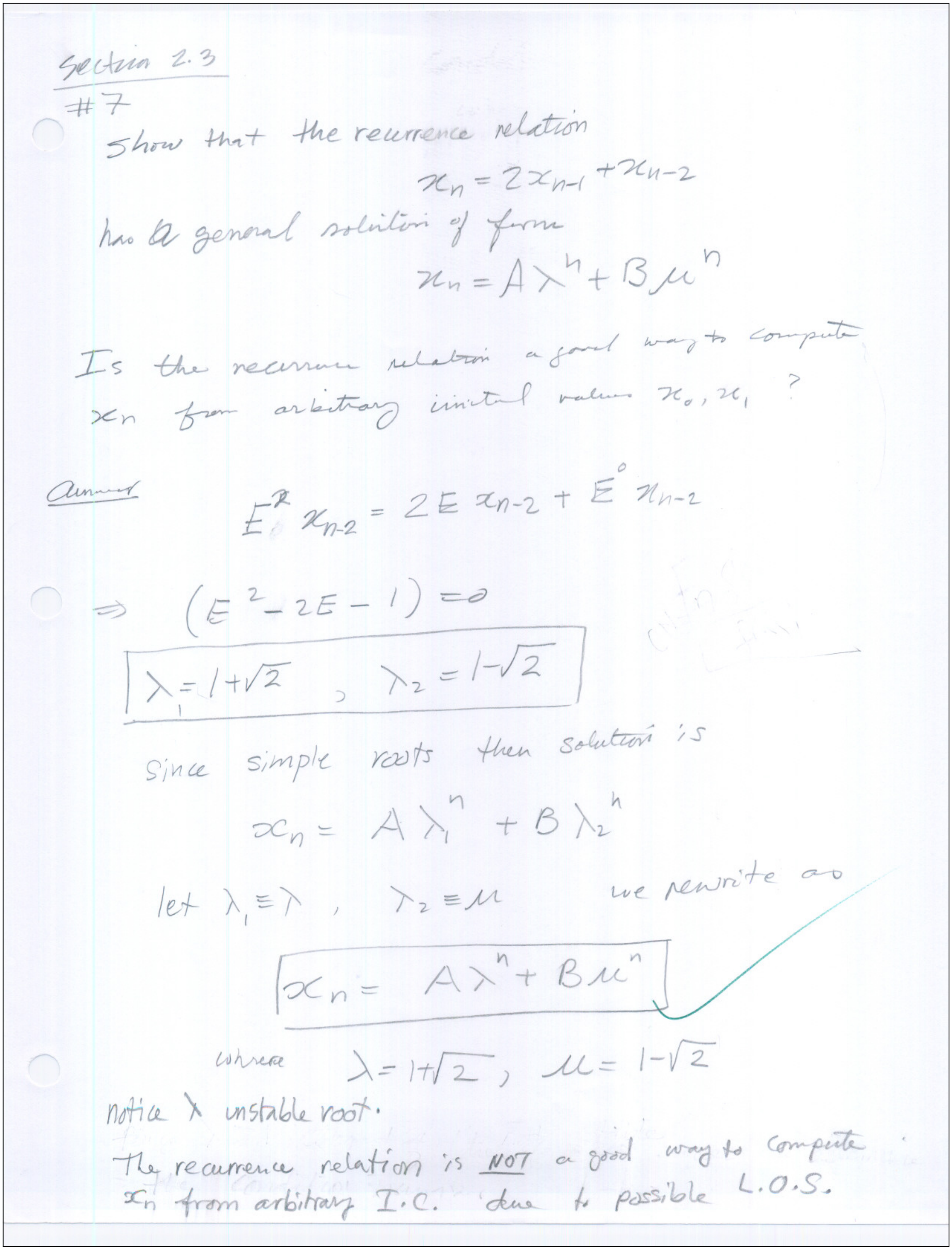
$$\boxed{CN \to \infty \quad \text{when} \quad x \to \infty}$$

(f) $\frac{1}{\cos x}$

$$CN = x\frac{f'(x)}{f(x)} = x \tan x$$

so at $\boxed{x = \pm n\frac{\pi}{2}}$ for $n = 1, 2, 3, \ldots$

$\boxed{CN \to \infty}$.



Section 2.3

#7

Show that the recurrence relation

$$x_n = 2x_{n-1} + x_{n-2}$$

has a general solution of form

$$x_n = A\lambda^n + B\mu^n$$

Is the recurrence relation a good way to compute x_n from arbitrary initial values x_0, x_1 ?

answer

$$E^2 x_{n-2} = 2E\, x_{n-2} + E^0 x_{n-2}$$

$$\Rightarrow \left(E^2 - 2E - 1\right) = 0$$

$$\boxed{\lambda_1 = 1+\sqrt{2} \quad , \quad \lambda_2 = 1-\sqrt{2}}$$

since simple roots then solution is

$$x_n = A\lambda_1^n + B\lambda_2^n$$

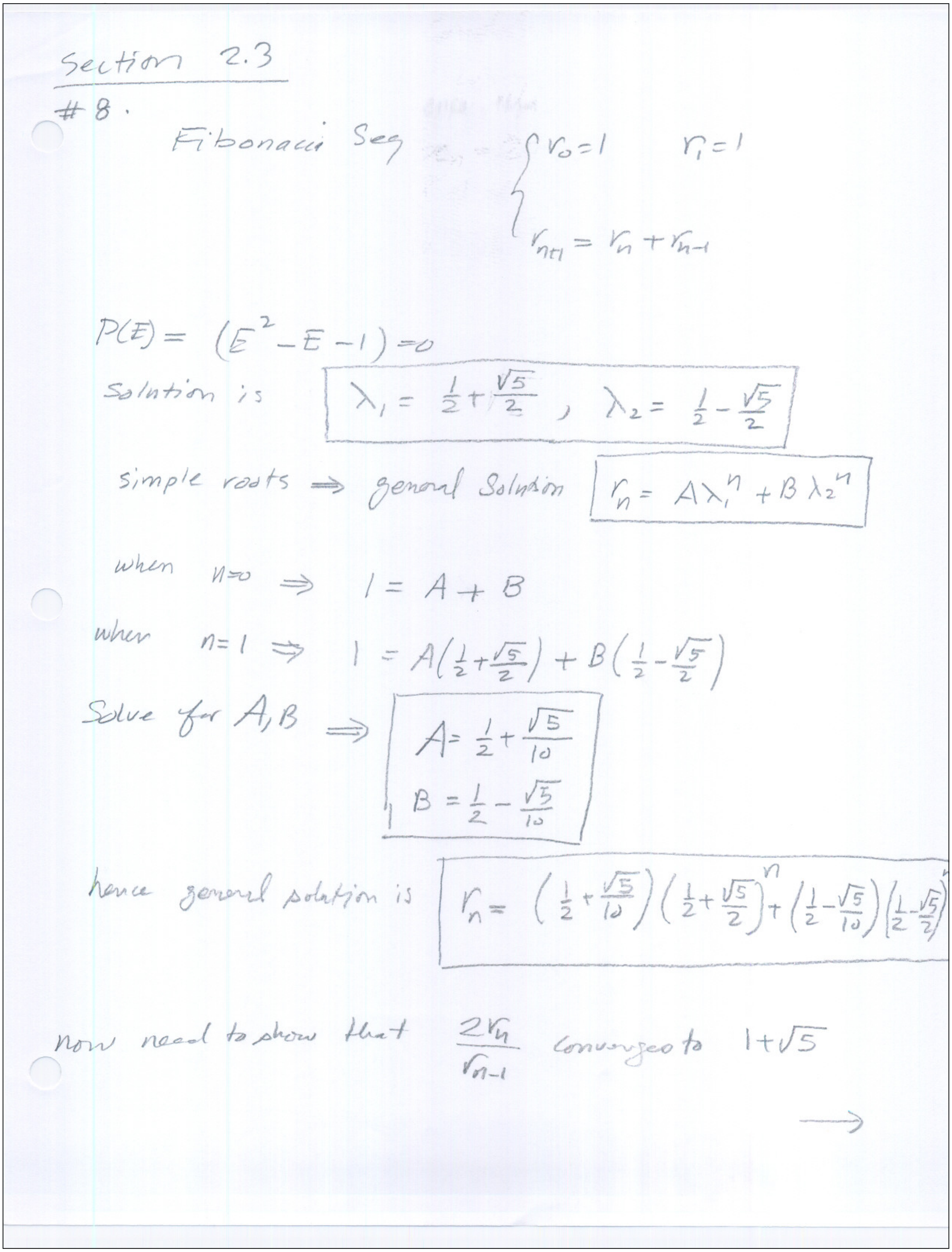
let $\lambda_1 \equiv \lambda$, $\lambda_2 \equiv \mu$ we rewrite as

$$\boxed{x_n = A\lambda^n + B\mu^n}$$

where $\lambda = 1+\sqrt{2}$, $\mu = 1-\sqrt{2}$

notice λ unstable root.

The recurrence relation is NOT a good way to compute x_n from arbitrary I.C. due to possible L.O.S.



Section 2.3

#8.

Fibonacci Seq
$$\begin{cases} r_0=1 \quad r_1=1 \\ r_{n+1}=r_n+r_{n-1} \end{cases}$$

$P(E)=(E^2-E-1)=0$

Solution is $\boxed{\lambda_1=\frac{1}{2}+\frac{\sqrt{5}}{2}, \quad \lambda_2=\frac{1}{2}-\frac{\sqrt{5}}{2}}$

simple roots $\Rightarrow$ general Solution $\boxed{r_n=A\lambda_1^n+B\lambda_2^n}$

when $n=0 \Rightarrow 1=A+B$

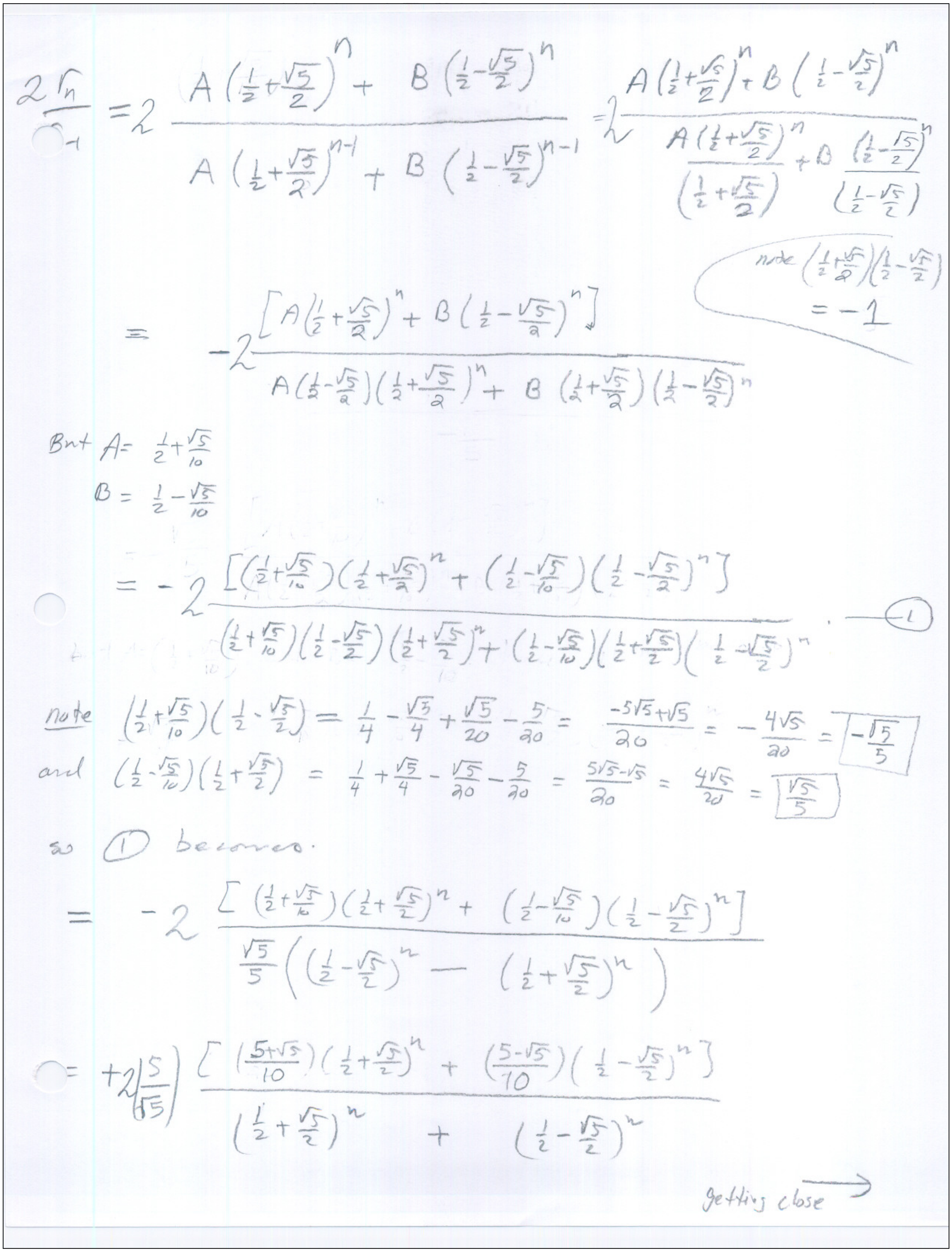
when $n=1 \Rightarrow 1=A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)+B\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)$

Solve for A,B $\Rightarrow$ $\boxed{\begin{aligned} A&=\frac{1}{2}+\frac{\sqrt{5}}{10} \\ B&=\frac{1}{2}-\frac{\sqrt{5}}{10} \end{aligned}}$

hence general solution is $\boxed{r_n=\left(\frac{1}{2}+\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n+\left(\frac{1}{2}-\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)}$

now need to show that $\frac{2r_n}{r_{n-1}}$ converges to $1+\sqrt{5}$

$\longrightarrow$



$$\frac{2r_n}{r_{n-1}} = 2\,\frac{A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + B\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n}{A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^{n-1} + B\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^{n-1}} = 2\,\frac{A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + B\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n}{\frac{A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n}{\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)} + B\,\frac{\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n}{\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)}}$$

note $\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right) = -1$

$$= -2\,\frac{\left[A\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + B\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n\right]}{A\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + B\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n}$$

But $A = \frac{1}{2}+\frac{\sqrt{5}}{10}$

$B = \frac{1}{2}-\frac{\sqrt{5}}{10}$

$$= -2\,\frac{\left[\left(\frac{1}{2}+\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + \left(\frac{1}{2}-\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n\right]}{\left(\frac{1}{2}+\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + \left(\frac{1}{2}-\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n} \quad (1)$$

note $\left(\frac{1}{2}+\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right) = \frac{1}{4}-\frac{\sqrt{5}}{4}+\frac{\sqrt{5}}{20}-\frac{5}{20} = \frac{-5\sqrt{5}+\sqrt{5}}{20} = -\frac{4\sqrt{5}}{20} = \boxed{-\frac{\sqrt{5}}{5}}$

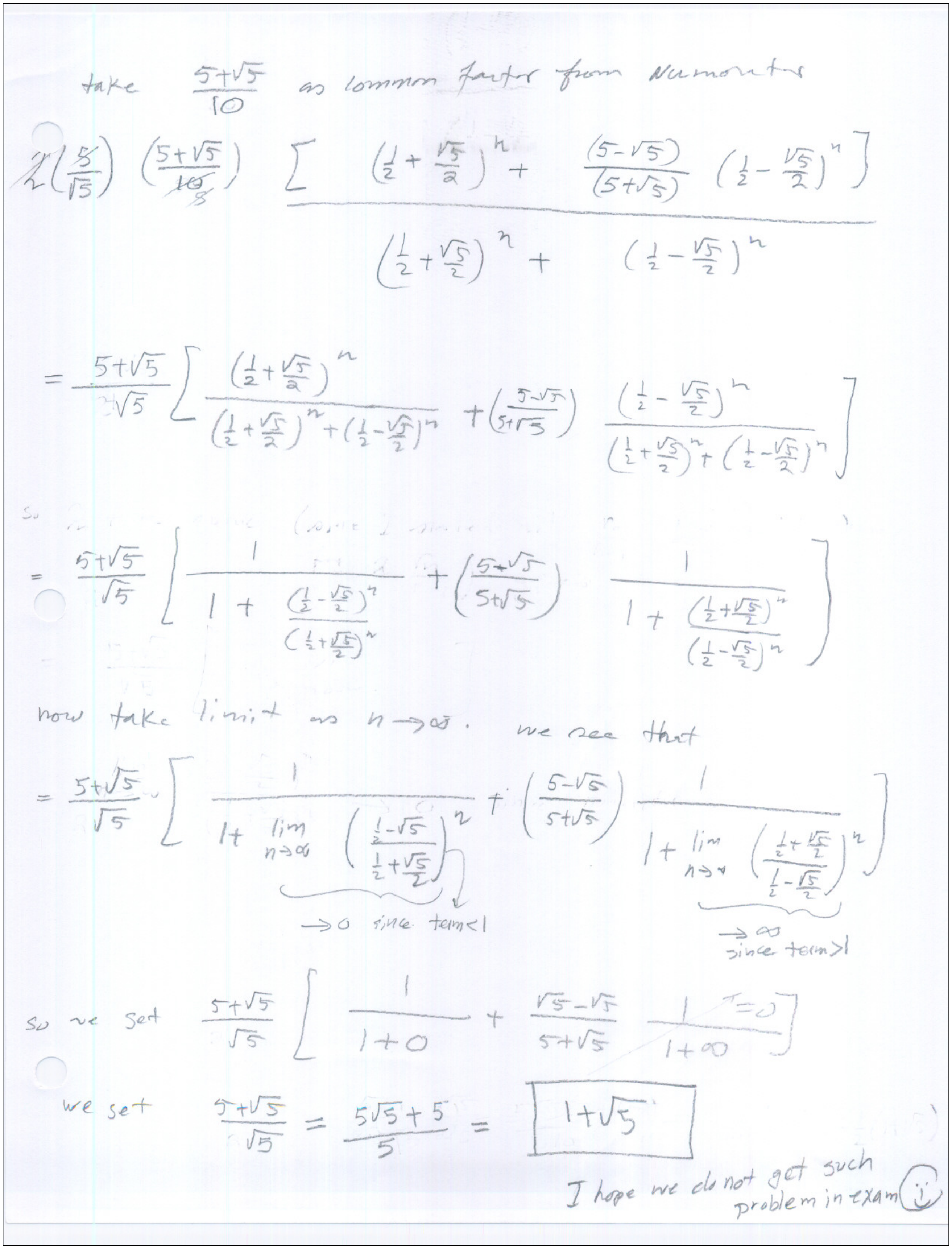
and $\left(\frac{1}{2}-\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right) = \frac{1}{4}+\frac{\sqrt{5}}{4}-\frac{\sqrt{5}}{20}-\frac{5}{20} = \frac{5\sqrt{5}-\sqrt{5}}{20} = \frac{4\sqrt{5}}{20} = \boxed{\frac{\sqrt{5}}{5}}$

so (1) becomes.

$$= -2\,\frac{\left[\left(\frac{1}{2}+\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + \left(\frac{1}{2}-\frac{\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n\right]}{\frac{\sqrt{5}}{5}\left(\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n - \left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n\right)}$$

$$= +2\left(\frac{5}{\sqrt{5}}\right)\frac{\left[\left(\frac{5+\sqrt{5}}{10}\right)\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + \left(\frac{5-\sqrt{5}}{10}\right)\left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n\right]}{\left(\frac{1}{2}+\frac{\sqrt{5}}{2}\right)^n + \left(\frac{1}{2}-\frac{\sqrt{5}}{2}\right)^n}$$

getting close →



take $\frac{5+\sqrt5}{10}$ as common factor from Numerator

$$\cancel{2}\left(\frac{5}{\sqrt5}\right)\left(\frac{5+\sqrt5}{\cancel{10}\,8}\right)\frac{\left[\left(\frac12+\frac{\sqrt5}{2}\right)^n+\frac{(5-\sqrt5)}{(5+\sqrt5)}\left(\frac12-\frac{\sqrt5}{2}\right)^n\right]}{\left(\frac12+\frac{\sqrt5}{2}\right)^n+\left(\frac12-\frac{\sqrt5}{2}\right)^n}$$

$$=\frac{5+\sqrt5}{2\sqrt5}\left[\frac{\left(\frac12+\frac{\sqrt5}{2}\right)^n}{\left(\frac12+\frac{\sqrt5}{2}\right)^n+\left(\frac12-\frac{\sqrt5}{2}\right)^n}+\left(\frac{5-\sqrt5}{5+\sqrt5}\right)\frac{\left(\frac12-\frac{\sqrt5}{2}\right)^n}{\left(\frac12+\frac{\sqrt5}{2}\right)^n+\left(\frac12-\frac{\sqrt5}{2}\right)^n}\right]$$

$$=\frac{5+\sqrt5}{\sqrt5}\left[\frac{1}{1+\frac{\left(\frac12-\frac{\sqrt5}{2}\right)^n}{\left(\frac12+\frac{\sqrt5}{2}\right)^n}}+\left(\frac{5-\sqrt5}{5+\sqrt5}\right)\frac{1}{1+\frac{\left(\frac12+\frac{\sqrt5}{2}\right)^n}{\left(\frac12-\frac{\sqrt5}{2}\right)^n}}\right]$$

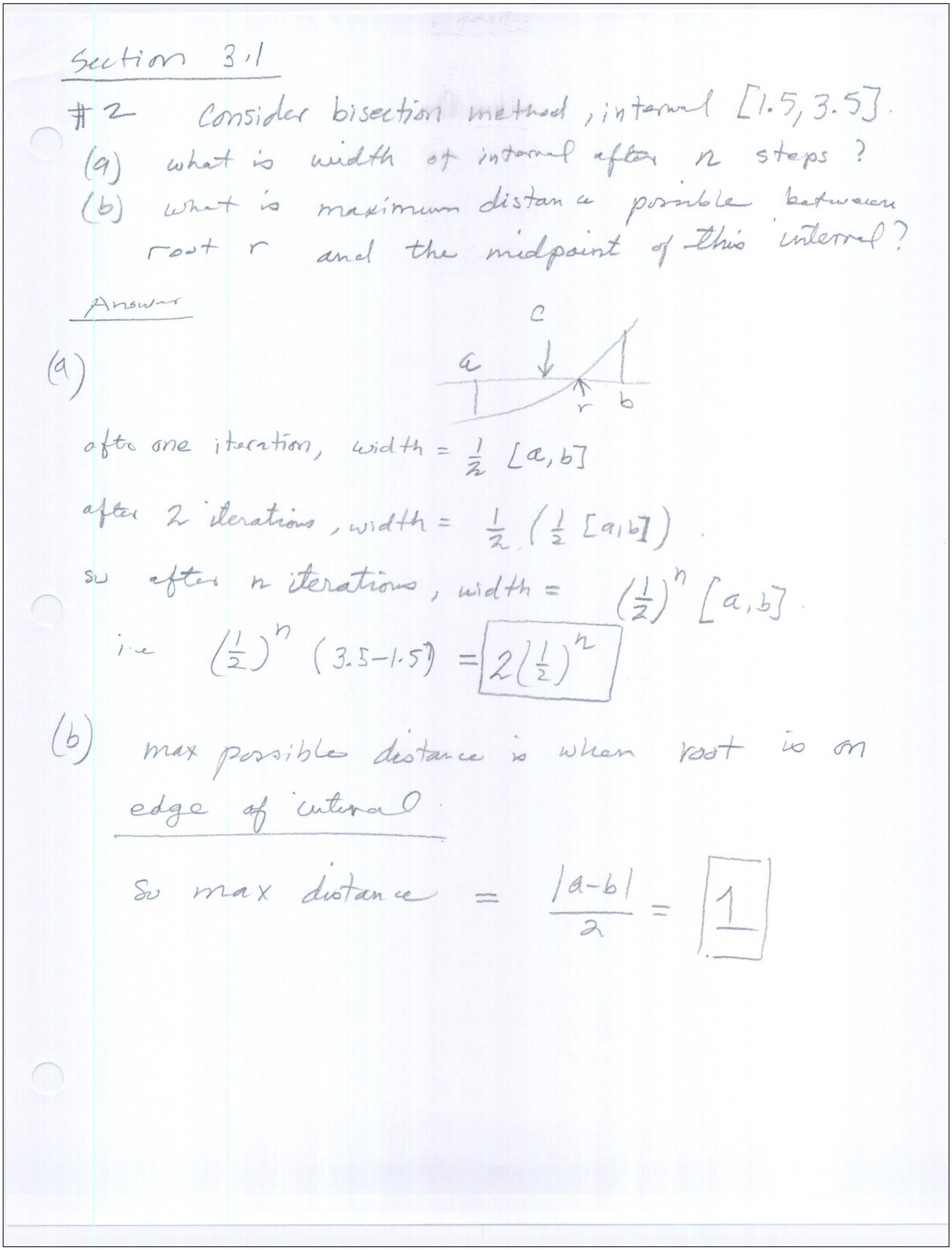
now take limit as $n\to\infty$. we see that

$$=\frac{5+\sqrt5}{\sqrt5}\left[\frac{1}{1+\underbrace{\lim_{n\to\infty}\left(\frac{\frac12-\frac{\sqrt5}{2}}{\frac12+\frac{\sqrt5}{2}}\right)^n}_{\to 0 \text{ since term}<1}}+\left(\frac{5-\sqrt5}{5+\sqrt5}\right)\frac{1}{1+\underbrace{\lim_{n\to\infty}\left(\frac{\frac12+\frac{\sqrt5}{2}}{\frac12-\frac{\sqrt5}{2}}\right)^n}_{\to\infty \text{ since term}>1}}\right]$$

So we set $\frac{5+\sqrt5}{\sqrt5}\left[\frac{1}{1+0}+\frac{\sqrt5-\sqrt5}{5+\sqrt5}\cancel{\frac{1}{1+\infty}}=0\right]$

we set $\frac{5+\sqrt5}{\sqrt5}=\frac{5\sqrt5+5}{5}=\boxed{1+\sqrt5}$

I hope we do not get such problem in exam ☺



Section 3.1

#2 Consider bisection method, interval [1.5, 3.5].

(a) what is width of interval after n steps?

(b) what is maximum distance possible between root r and the midpoint of this interval?

Answer

(a)

after one iteration, width = $\frac{1}{2}[a,b]$

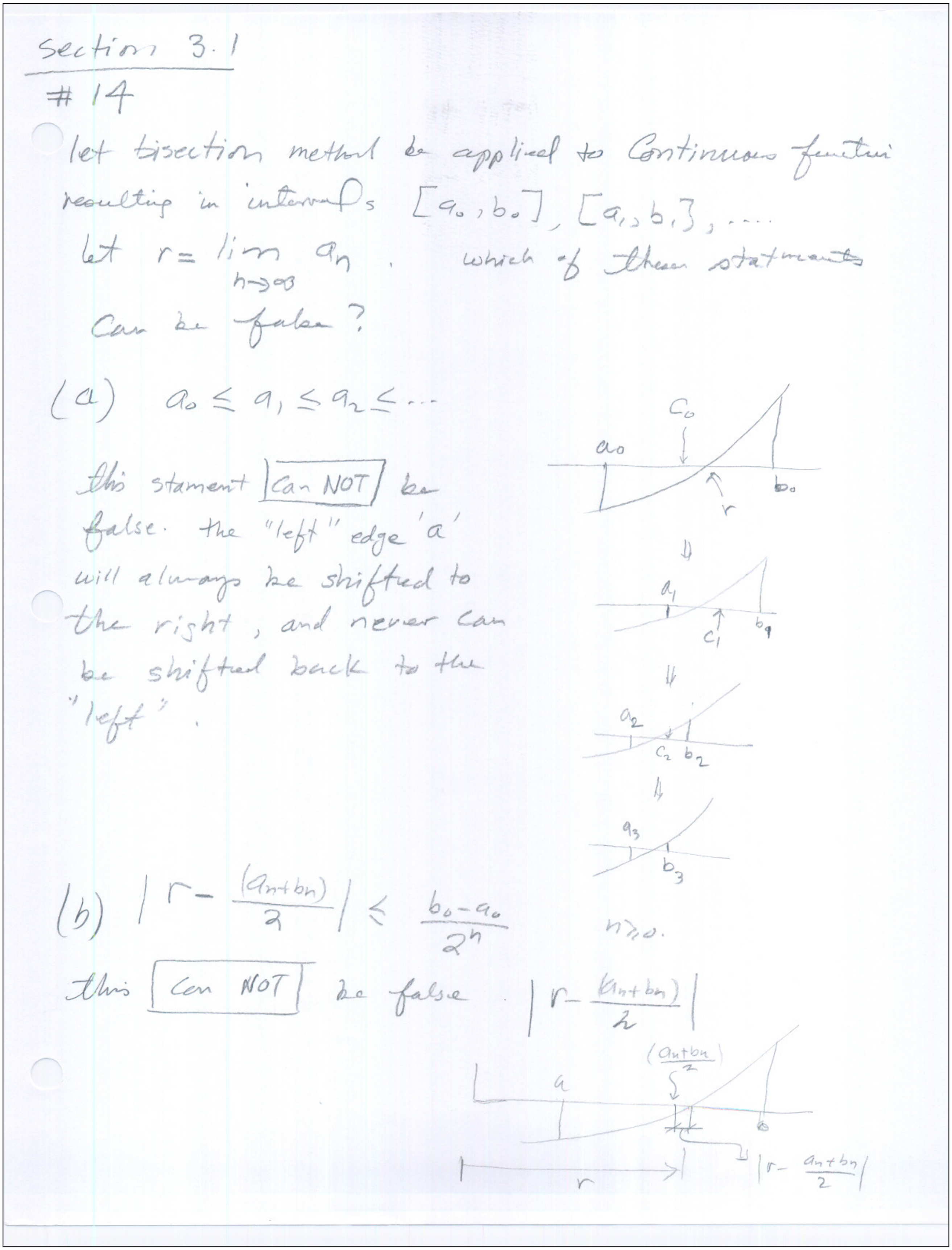
after 2 iterations, width = $\frac{1}{2}\left(\frac{1}{2}[a,b]\right)$.

so after n iterations, width = $\left(\frac{1}{2}\right)^n [a,b]$.

i.e $\left(\frac{1}{2}\right)^n (3.5-1.5) = \boxed{2\left(\frac{1}{2}\right)^n}$

(b) max possible distance is when root is on edge of interval.

So max distance $= \frac{|a-b|}{2} = \boxed{1}$



section 3.1

#14

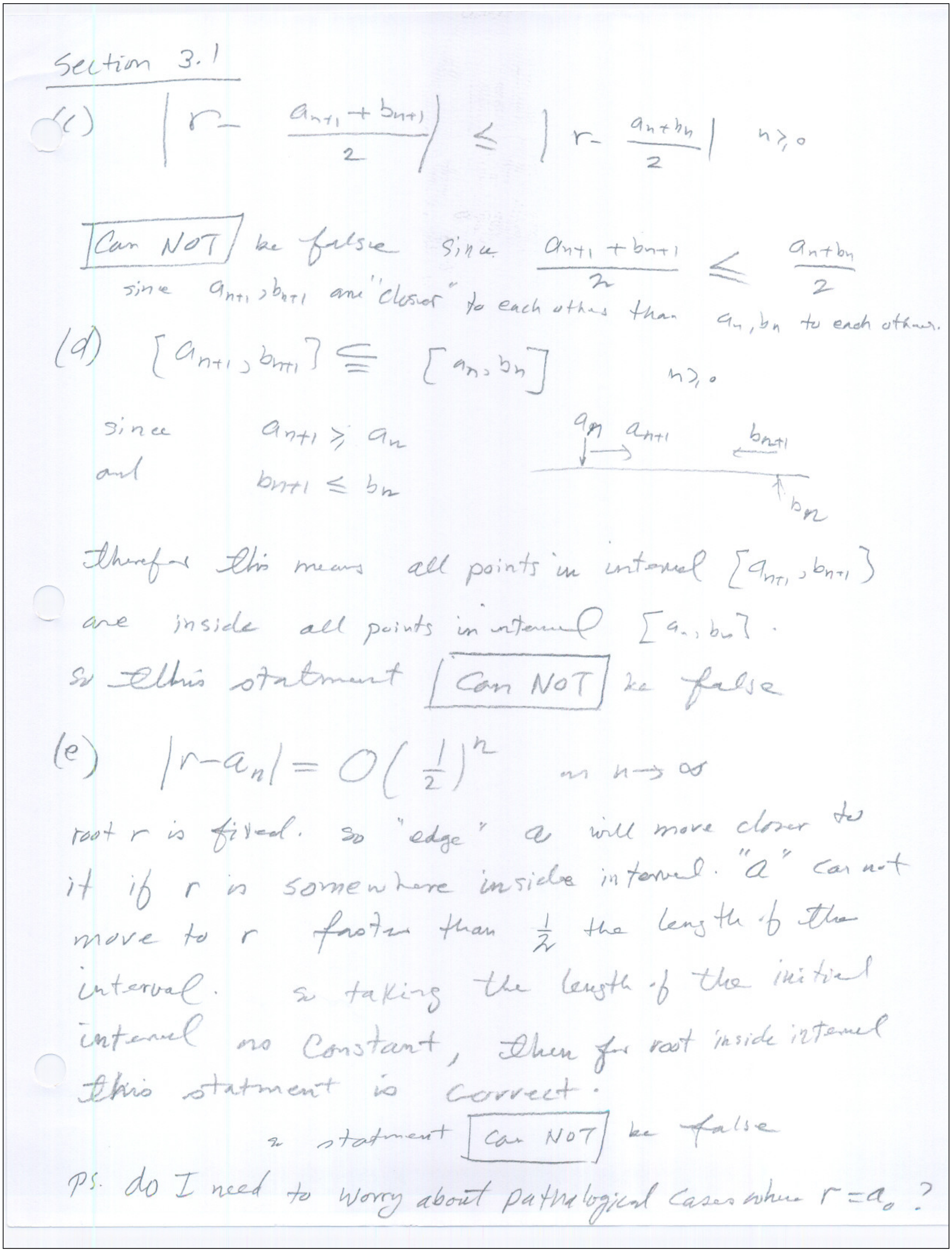
let bisection method be applied to Continuous function resulting in intervals $[a_0, b_0], [a_1, b_1], \ldots$ let $r = \lim_{n\to\infty} a_n$. which of these statments can be false?

(a) $a_0 \le a_1 \le a_2 \le \ldots$

this stament $\boxed{\text{can NOT}}$ be false. the "left" edge 'a' will always be shifted to the right, and never can be shifted back to the "left".

(b) $\left| r - \frac{(a_n+b_n)}{2} \right| \le \frac{b_0-a_0}{2^n}$ $n \ge 0$.

this $\boxed{\text{Can NOT}}$ be false. $\left| r - \frac{(a_n+b_n)}{2} \right|$



Section 3.1

(c) $\left| r - \frac{a_{n+1}+b_{n+1}}{2} \right| \le \left| r - \frac{a_n+b_n}{2} \right|$ $n \geq 0$

$\boxed{\text{Can NOT}}$ be false since $\frac{a_{n+1}+b_{n+1}}{2} \le \frac{a_n+b_n}{2}$

since a_{n+1}, b_{n+1} are "closer" to each other than a_n, b_n to each other.

(d) $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$ $n \geq 0$

since $a_{n+1} \geq a_n$

and $b_{n+1} \le b_n$

a_n
a_{n+1}
b_{n+1}
b_n

therefore this means all points in interval $\{a_{n+1}, b_{n+1}\}$ are inside all points in interval $[a_n, b_n]$.

so this statment $\boxed{\text{Can NOT}}$ be false

(e) $|r-a_n| = O\left(\frac{1}{2}\right)^n$ as $n \to \infty$

root r is fixed. so "edge" a will move closer to it if r is somewhere inside interval. "a" can not move to r faster than $\frac{1}{2}$ the length of the interval. so taking the length of the initial interval as constant, then for root inside interval this statment is correct.

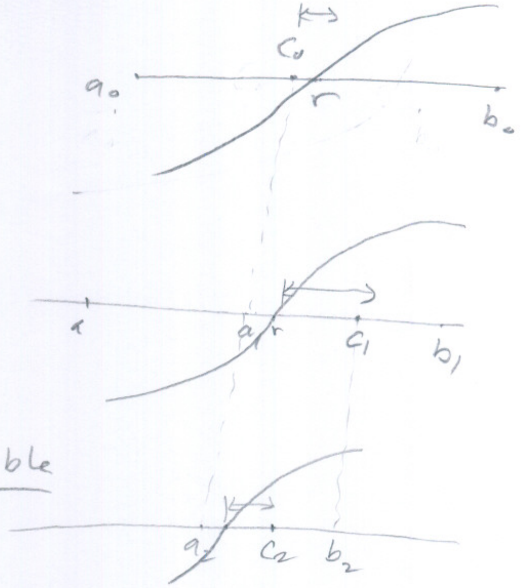
so statment $\boxed{\text{can NOT}}$ be false

PS. do I need to worry about pathalogical cases when $r = a_0$?

section 3.1

(f) $|r - c_n| < |r - c_{n-1}|$ $n \geq 1$

c_n is the middle of the interval at step n.



We see that it is NOT possible that $|r - c_n| > |r - c_{n-1}|$

So this statment Can NOT be false

ps. for n=0, it is possible that

$|r - c_0| < |r - c_1|$ but for $n \geq 1$ Not possible.

Section 3.1 # 15

prove that point c computed in bisection is the point where line through

(a, sign(f(a)) and (b, sign(f(b)))

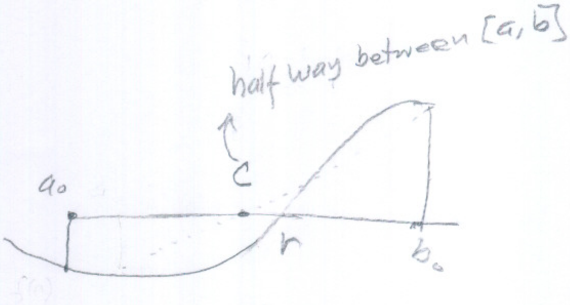
intersects x– axis.

Note

$\text{Sign}(f(a))=1$ if $f(a)>0$

$\text{Sign}(f(a))=-1$ if $f(a)<0$.

$\text{sign}(f(a))=0$ if $f(a)=0$



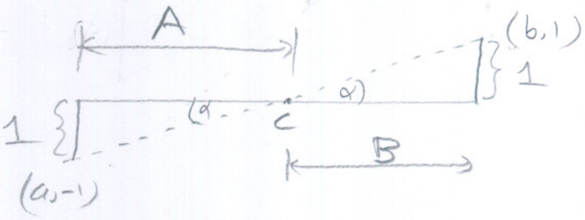
⇓

Assume $f(a)<0$ and $f(b)>0$.

then the line is $(a,-1)$ and $(b,1)$ as follows

need to show that

$A=B$.

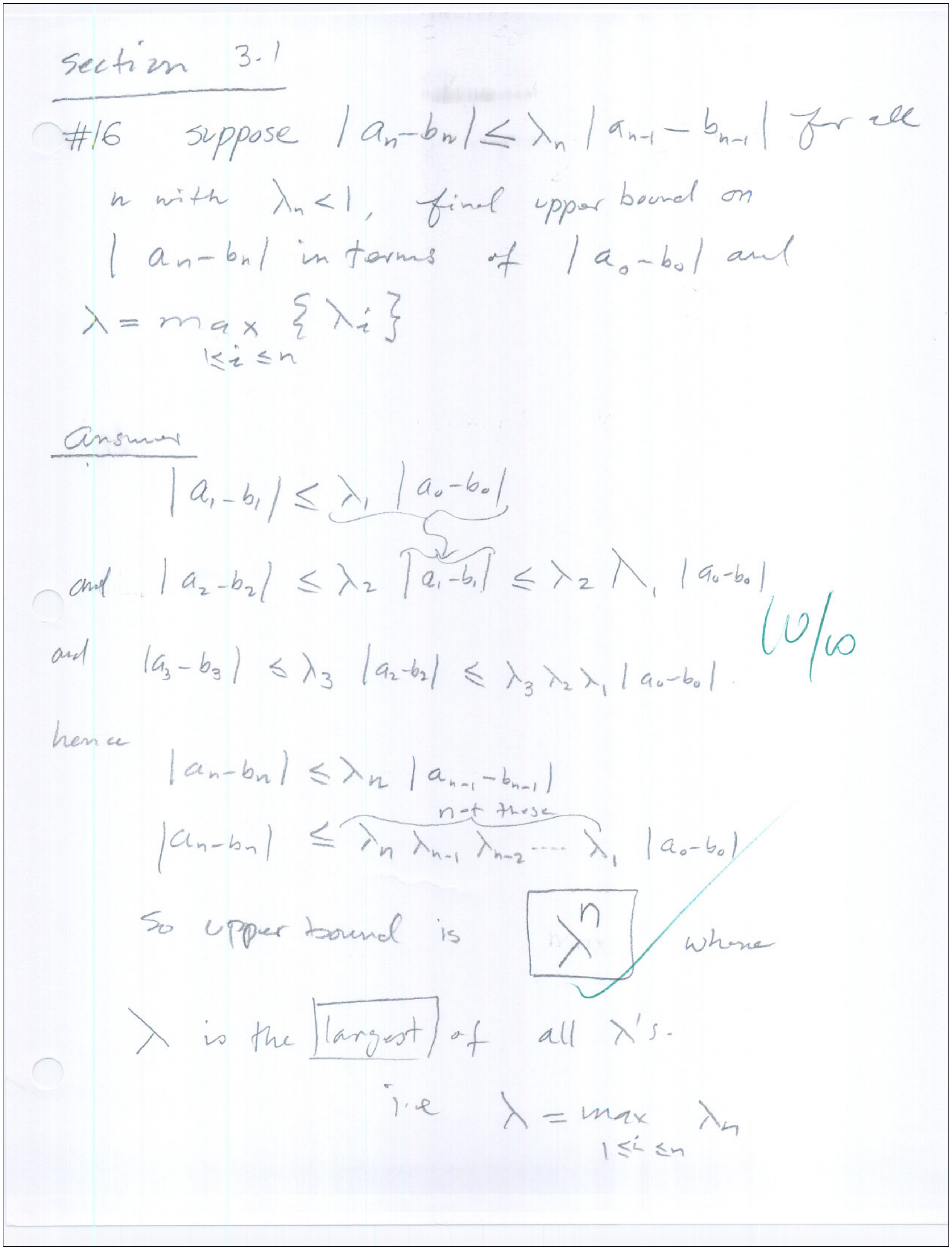


since $\tan\alpha=\frac{1}{B}$

and also $\tan\alpha=\frac{1}{A}$

hence $\frac{1}{B}=\frac{1}{A} \Rightarrow B=A$ hence line intersect

x– axis at mid point, which is point C. ✓



section 3.1

#16 suppose $|a_n-b_n| \le \lambda_n |a_{n-1}-b_{n-1}|$ for all n with $\lambda_n<1$, final upper bound on $|a_n-b_n|$ in terms of $|a_0-b_0|$ and $\lambda = \max\limits_{1\le i\le n}\{\lambda_i\}$

answer

$$|a_1-b_1| \le \underbrace{\lambda_1 |a_0-b_0|}$$

and $|a_2-b_2| \le \lambda_2 |a_1-b_1| \le \lambda_2 \lambda_1 |a_0-b_0|$

10/10

and $|a_3-b_3| \le \lambda_3 |a_2-b_2| \le \lambda_3\lambda_2\lambda_1 |a_0-b_0|$.

hence

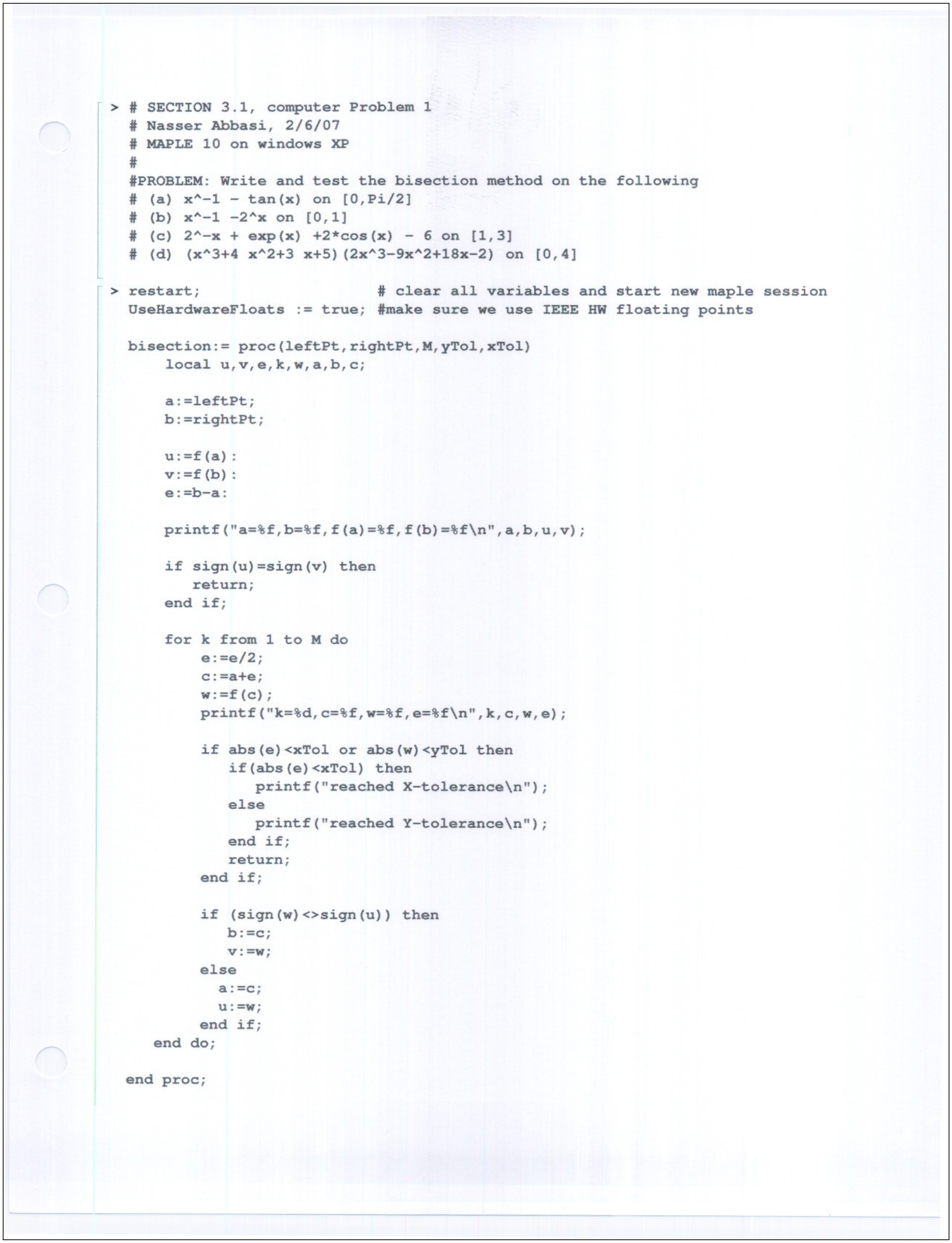
$$|a_n-b_n| \le \lambda_n |a_{n-1}-b_{n-1}|$$

$$|a_n-b_n| \le \overbrace{\lambda_n \lambda_{n-1} \lambda_{n-2} \cdots \lambda_1}^{n \text{ of these}} |a_0-b_0|$$

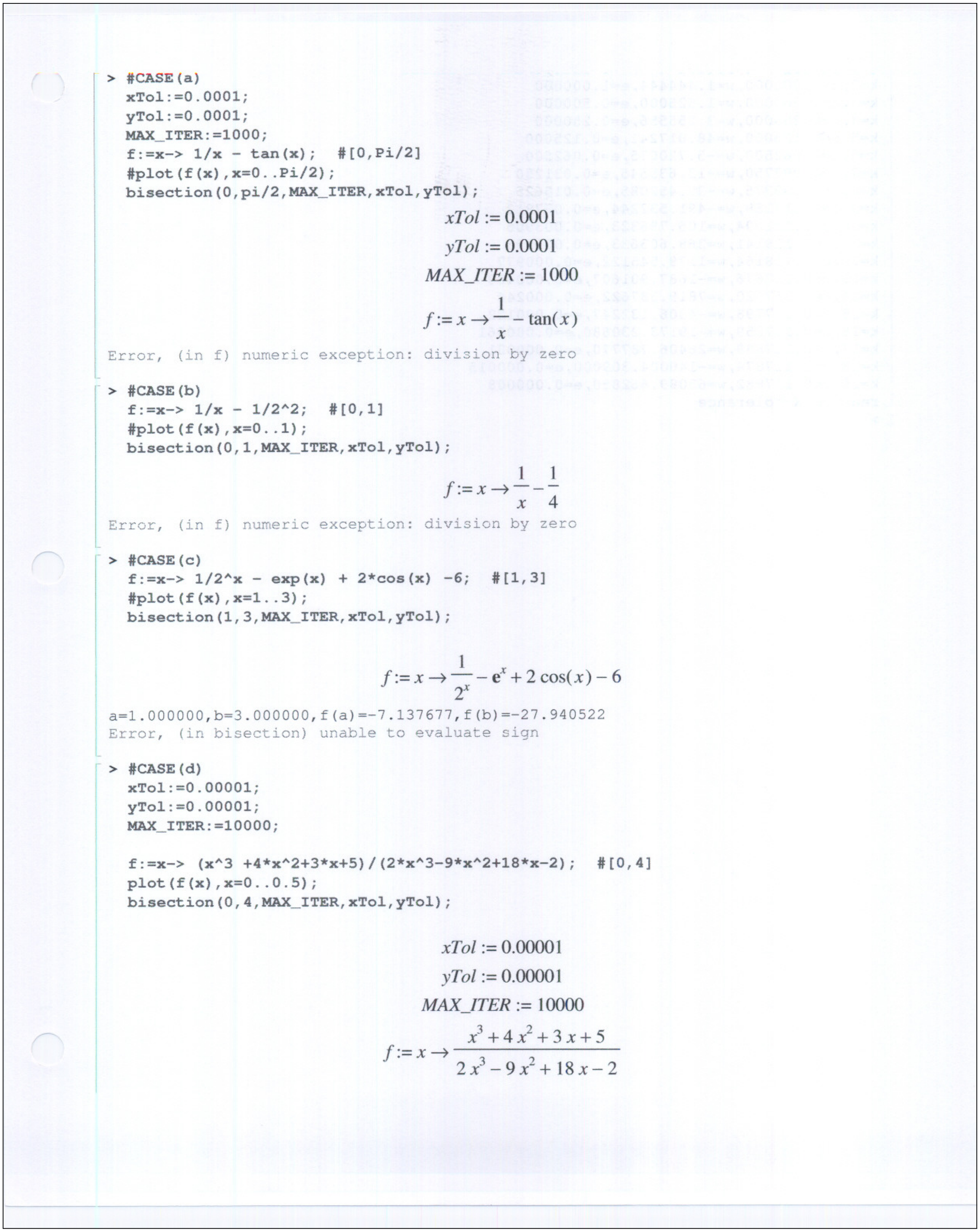
So upper bound is $\boxed{\lambda^n}$ where

λ is the $\boxed{\text{largest}}$ of all λ's.

i.e $\lambda = \max\limits_{1\le i\le n} \lambda_n$



```
> # SECTION 3.1, computer Problem 1
  # Nasser Abbasi, 2/6/07
  # MAPLE 10 on windows XP
  #
  #PROBLEM: Write and test the bisection method on the following
  # (a) x^-1 - tan(x) on [0,Pi/2]
  # (b) x^-1 -2^x on [0,1]
  # (c) 2^-x + exp(x) +2*cos(x) - 6 on [1,3]
  # (d) (x^3+4 x^2+3 x+5)(2x^3-9x^2+18x-2) on [0,4]

> restart;                   # clear all variables and start new maple session
  UseHardwareFloats := true; #make sure we use IEEE HW floating points

  bisection:= proc(leftPt,rightPt,M,yTol,xTol)
      local u,v,e,k,w,a,b,c;

      a:=leftPt;
      b:=rightPt;

      u:=f(a):
      v:=f(b):
      e:=b-a:

      printf("a=%f,b=%f,f(a)=%f,f(b)=%f\n",a,b,u,v);

      if sign(u)=sign(v) then
         return;
      end if;

      for k from 1 to M do
          e:=e/2;
          c:=a+e;
          w:=f(c);
          printf("k=%d,c=%f,w=%f,e=%f\n",k,c,w,e);

          if abs(e)<xTol or abs(w)<yTol then
             if(abs(e)<xTol) then
                printf("reached X-tolerance\n");
             else
                printf("reached Y-tolerance\n");
             end if;
             return;
          end if;

          if (sign(w)<>sign(u)) then
             b:=c;
             v:=w;
          else
            a:=c;
            u:=w;
          end if;
    end do;

  end proc;
```



```
> #CASE(a)
  xTol:=0.0001;
  yTol:=0.0001;
  MAX_ITER:=1000;
  f:=x-> 1/x - tan(x);   #[0,Pi/2]
  #plot(f(x),x=0..Pi/2);
  bisection(0,pi/2,MAX_ITER,xTol,yTol);
```

$$xTol := 0.0001$$

$$yTol := 0.0001$$

$$MAX\_ITER := 1000$$

$$f := x \rightarrow \frac{1}{x} - \tan(x)$$

```
Error, (in f) numeric exception: division by zero
```

```
> #CASE(b)
  f:=x-> 1/x - 1/2^2;   #[0,1]
  #plot(f(x),x=0..1);
  bisection(0,1,MAX_ITER,xTol,yTol);
```

$$f := x \rightarrow \frac{1}{x} - \frac{1}{4}$$

```
Error, (in f) numeric exception: division by zero
```

```
> #CASE(c)
  f:=x-> 1/2^x - exp(x) + 2*cos(x) -6;   #[1,3]
  #plot(f(x),x=1..3);
  bisection(1,3,MAX_ITER,xTol,yTol);
```

$$f := x \rightarrow \frac{1}{2^x} - \mathrm{e}^x + 2\cos(x) - 6$$

```
a=1.000000,b=3.000000,f(a)=-7.137677,f(b)=-27.940522
Error, (in bisection) unable to evaluate sign
```

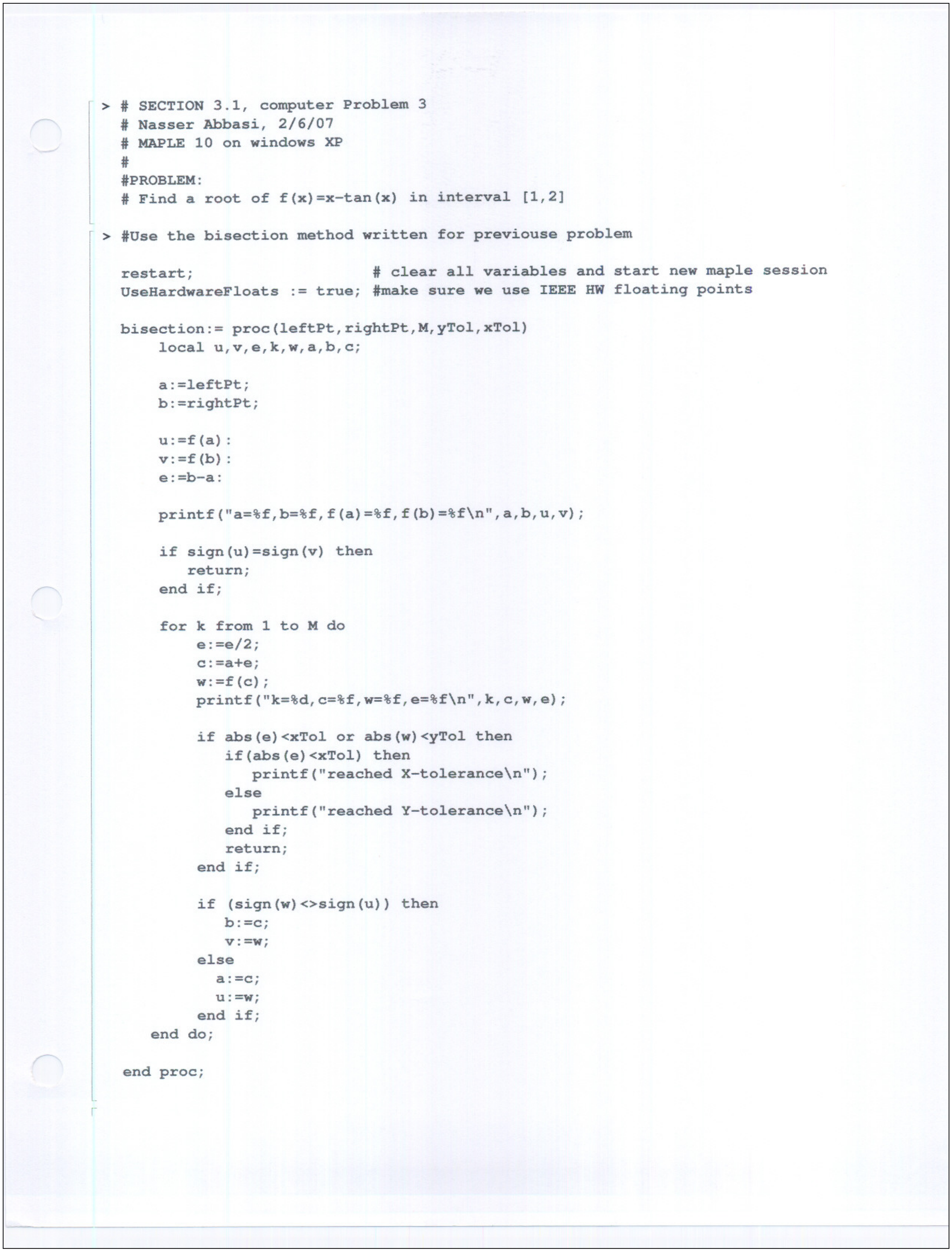
```
> #CASE(d)
  xTol:=0.00001;
  yTol:=0.00001;
  MAX_ITER:=10000;

  f:=x-> (x^3 +4*x^2+3*x+5)/(2*x^3-9*x^2+18*x-2);   #[0,4]
  plot(f(x),x=0..0.5);
  bisection(0,4,MAX_ITER,xTol,yTol);
```

$$xTol := 0.00001$$

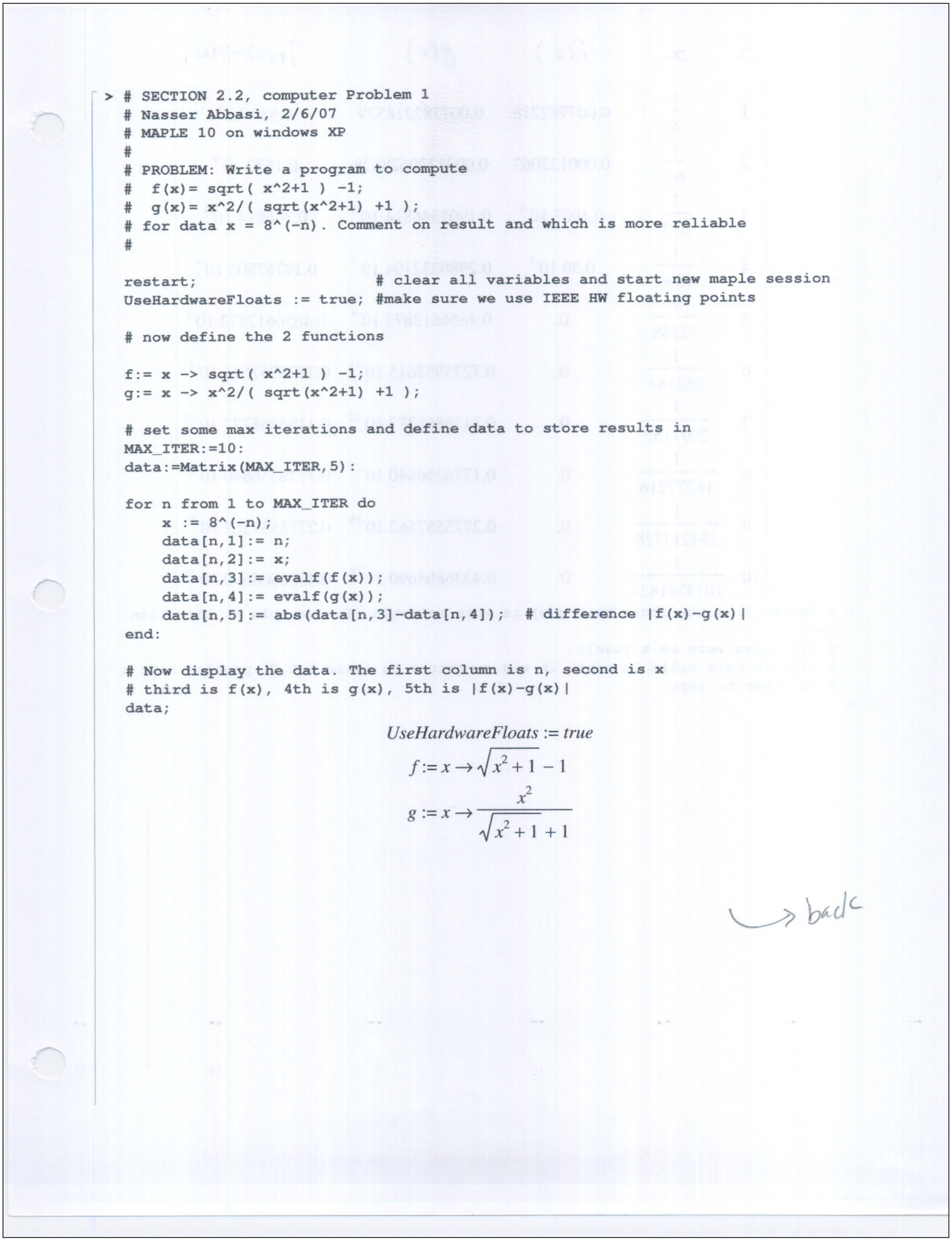
$$yTol := 0.00001$$

$$MAX\_ITER := 10000$$

$$f := x \rightarrow \frac{x^3 + 4x^2 + 3x + 5}{2x^3 - 9x^2 + 18x - 2}$$



```
> # SECTION 3.1, computer Problem 3
  # Nasser Abbasi, 2/6/07
  # MAPLE 10 on windows XP
  #
  #PROBLEM:
  # Find a root of f(x)=x-tan(x) in interval [1,2]

> #Use the bisection method written for previouse problem

  restart;                    # clear all variables and start new maple session
  UseHardwareFloats := true;  #make sure we use IEEE HW floating points

  bisection:= proc(leftPt,rightPt,M,yTol,xTol)
      local u,v,e,k,w,a,b,c;

      a:=leftPt;
      b:=rightPt;

      u:=f(a):
      v:=f(b):
      e:=b-a:

      printf("a=%f,b=%f,f(a)=%f,f(b)=%f\n",a,b,u,v);

      if sign(u)=sign(v) then
         return;
      end if;

      for k from 1 to M do
          e:=e/2;
          c:=a+e;
          w:=f(c);
          printf("k=%d,c=%f,w=%f,e=%f\n",k,c,w,e);

          if abs(e)<xTol or abs(w)<yTol then
             if(abs(e)<xTol) then
                printf("reached X-tolerance\n");
             else
                printf("reached Y-tolerance\n");
             end if;
             return;
          end if;

          if (sign(w)<>sign(u)) then
             b:=c;
             v:=w;
          else
            a:=c;
            u:=w;
          end if;
     end do;

  end proc;
```



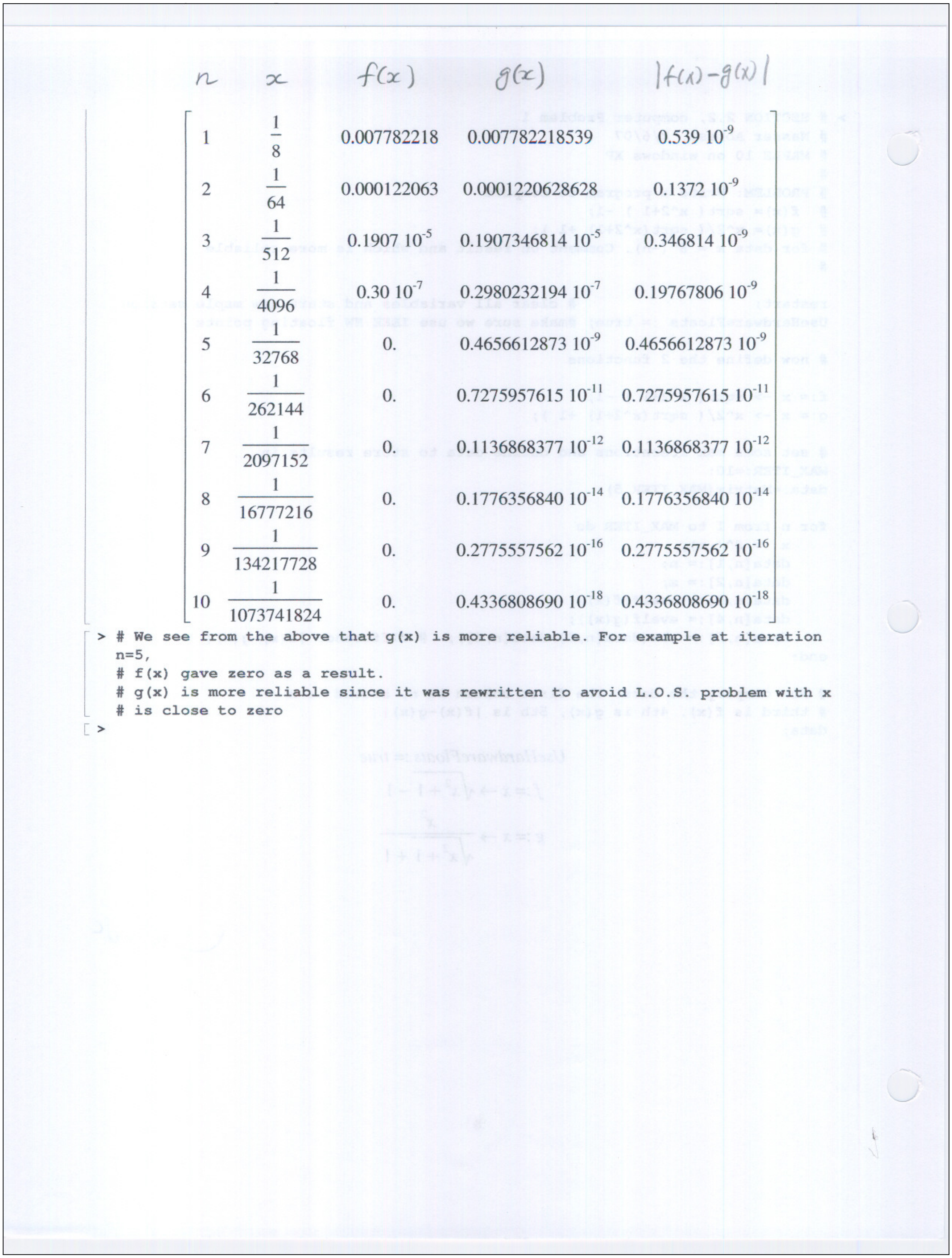
```
> # SECTION 2.2, computer Problem 1
  # Nasser Abbasi, 2/6/07
  # MAPLE 10 on windows XP
  #
  # PROBLEM: Write a program to compute
  #   f(x)= sqrt( x^2+1 ) -1;
  #   g(x)= x^2/( sqrt(x^2+1) +1 );
  # for data x = 8^(-n). Comment on result and which is more reliable
  #

  restart;                       # clear all variables and start new maple session
  UseHardwareFloats := true; #make sure we use IEEE HW floating points

  # now define the 2 functions

  f:= x -> sqrt( x^2+1 ) -1;
  g:= x -> x^2/( sqrt(x^2+1) +1 );

  # set some max iterations and define data to store results in
  MAX_ITER:=10:
  data:=Matrix(MAX_ITER,5):

  for n from 1 to MAX_ITER do
      x := 8^(-n);
      data[n,1]:= n;
      data[n,2]:= x;
      data[n,3]:= evalf(f(x));
      data[n,4]:= evalf(g(x));
      data[n,5]:= abs(data[n,3]-data[n,4]);  # difference |f(x)-g(x)|
  end:

  # Now display the data. The first column is n, second is x
  # third is f(x), 4th is g(x), 5th is |f(x)-g(x)|
  data;
```

$$UseHardwareFloats := true$$

$$f := x \rightarrow \sqrt{x^2+1} - 1$$

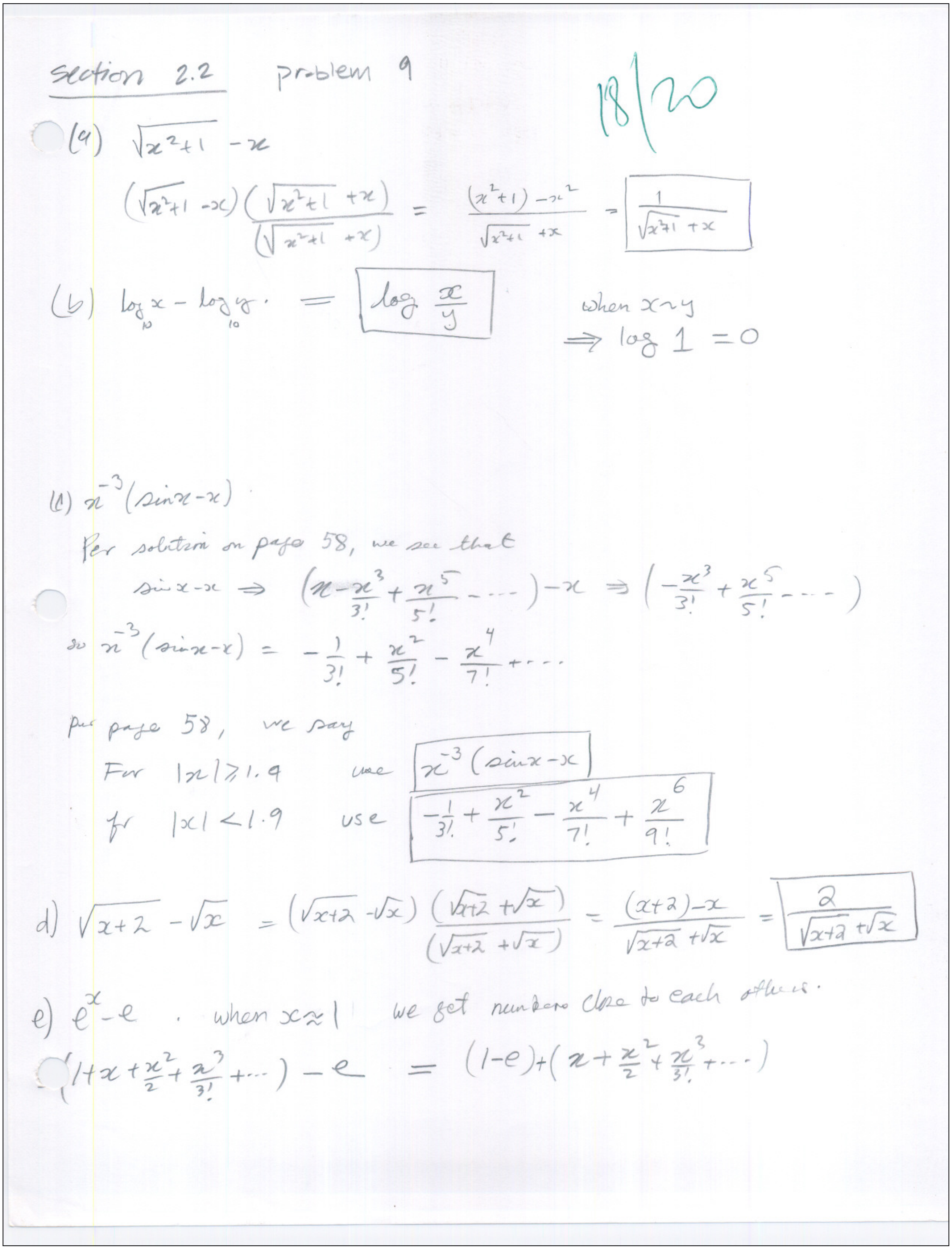
$$g := x \rightarrow \frac{x^2}{\sqrt{x^2+1}+1}$$

→ back



$$\begin{array}{ccccc}
n & x & f(x) & g(x) & |f(x)-g(x)| \\
1 & \frac{1}{8} & 0.007782218 & 0.007782218539 & 0.539\ 10^{-9} \\
2 & \frac{1}{64} & 0.000122063 & 0.0001220628628 & 0.1372\ 10^{-9} \\
3 & \frac{1}{512} & 0.1907\ 10^{-5} & 0.1907346814\ 10^{-5} & 0.346814\ 10^{-9} \\
4 & \frac{1}{4096} & 0.30\ 10^{-7} & 0.2980232194\ 10^{-7} & 0.19767806\ 10^{-9} \\
5 & \frac{1}{32768} & 0. & 0.4656612873\ 10^{-9} & 0.4656612873\ 10^{-9} \\
6 & \frac{1}{262144} & 0. & 0.7275957615\ 10^{-11} & 0.7275957615\ 10^{-11} \\
7 & \frac{1}{2097152} & 0. & 0.1136868377\ 10^{-12} & 0.1136868377\ 10^{-12} \\
8 & \frac{1}{16777216} & 0. & 0.1776356840\ 10^{-14} & 0.1776356840\ 10^{-14} \\
9 & \frac{1}{134217728} & 0. & 0.2775557562\ 10^{-16} & 0.2775557562\ 10^{-16} \\
10 & \frac{1}{1073741824} & 0. & 0.4336808690\ 10^{-18} & 0.4336808690\ 10^{-18}
\end{array}$$

```
> # We see from the above that g(x) is more reliable. For example at iteration
  n=5,
  # f(x) gave zero as a result.
  # g(x) is more reliable since it was rewritten to avoid L.O.S. problem with x
  # is close to zero
> 
```



section 2.2 problem 9

18/20

(a) $\sqrt{x^2+1}-x$

$$\frac{(\sqrt{x^2+1}-x)(\sqrt{x^2+1}+x)}{(\sqrt{x^2+1}+x)} = \frac{(x^2+1)-x^2}{\sqrt{x^2+1}+x} = \boxed{\frac{1}{\sqrt{x^2+1}+x}}$$

(b) $\log_{10} x - \log_{10} y. = \boxed{\log\frac{x}{y}}$ when $x\sim y$ $\Rightarrow \log 1 = 0$

(c) $x^{-3}(\sin x - x)$

Per solution on page 58, we see that

$$\sin x - x \Rightarrow \left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots\right)-x \Rightarrow \left(-\frac{x^3}{3!}+\frac{x^5}{5!}-\cdots\right)$$

so $x^{-3}(\sin x - x) = -\frac{1}{3!}+\frac{x^2}{5!}-\frac{x^4}{7!}+\cdots$

per page 58, we say

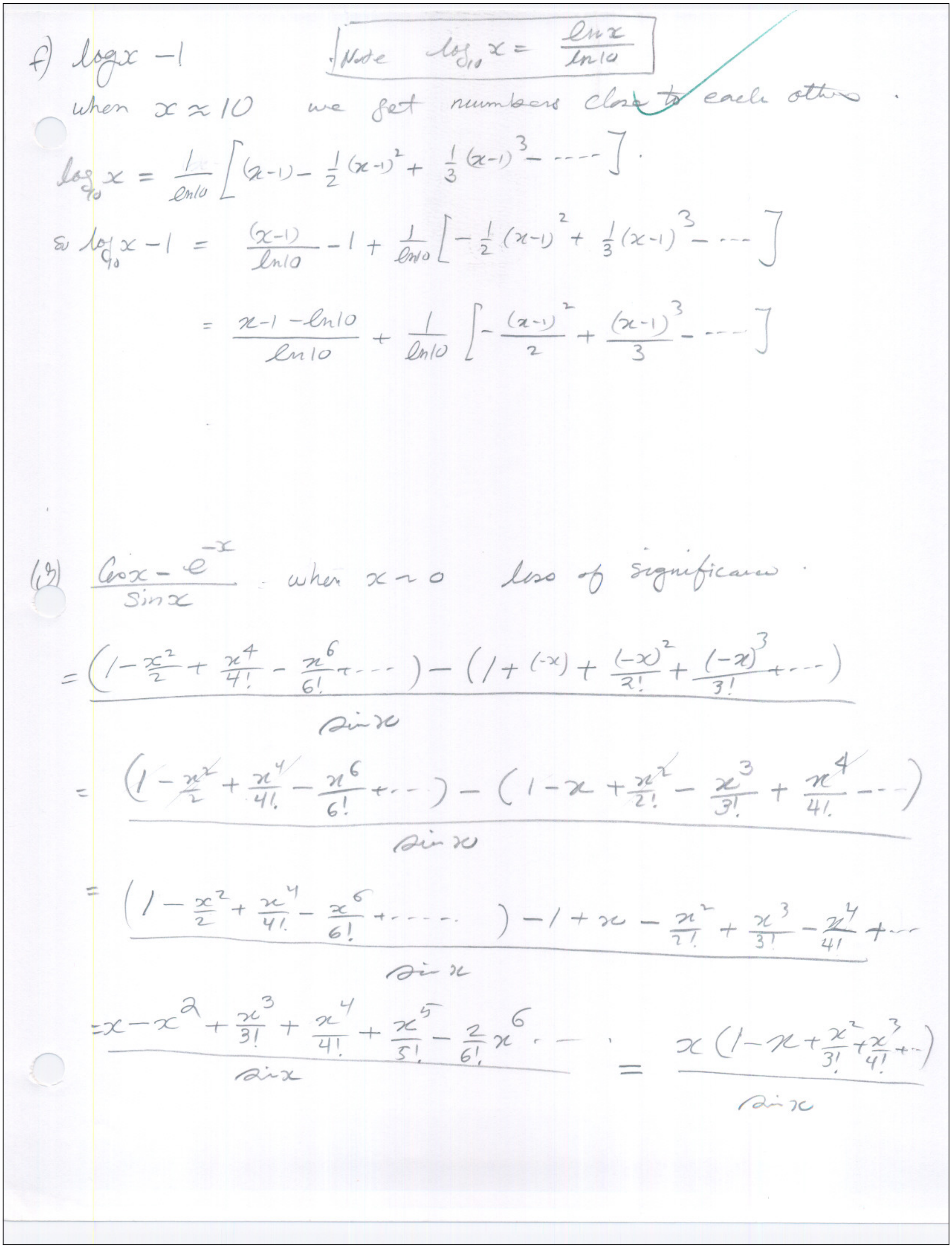
For $|x|\geq 1.9$ use $\boxed{x^{-3}(\sin x - x)}$

for $|x|<1.9$ use $\boxed{-\frac{1}{3!}+\frac{x^2}{5!}-\frac{x^4}{7!}+\frac{x^6}{9!}}$

d) $\sqrt{x+2}-\sqrt{x} = (\sqrt{x+2}-\sqrt{x})\frac{(\sqrt{x+2}+\sqrt{x})}{(\sqrt{x+2}+\sqrt{x})} = \frac{(x+2)-x}{\sqrt{x+2}+\sqrt{x}} = \boxed{\frac{2}{\sqrt{x+2}+\sqrt{x}}}$

e) e^x-e . when $x\approx 1$ we get numbers close to each others.

$$\left(1+x+\frac{x^2}{2}+\frac{x^3}{3!}+\cdots\right)-e = (1-e)+\left(x+\frac{x^2}{2}+\frac{x^3}{3!}+\cdots\right)$$



f) $\log x - 1$

$\boxed{\text{Note } \log_{10} x = \frac{\ln x}{\ln 10}}$

when $x \approx 10$ we get numbers close to each other.

$$\log_{10} x = \frac{1}{\ln 10}\left[(x-1) - \frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \cdots\right].$$

$$\text{so } \log_{10} x - 1 = \frac{(x-1)}{\ln 10} - 1 + \frac{1}{\ln 10}\left[-\frac{1}{2}(x-1)^2 + \frac{1}{3}(x-1)^3 - \cdots\right]$$

$$= \frac{x-1-\ln 10}{\ln 10} + \frac{1}{\ln 10}\left[-\frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots\right]$$

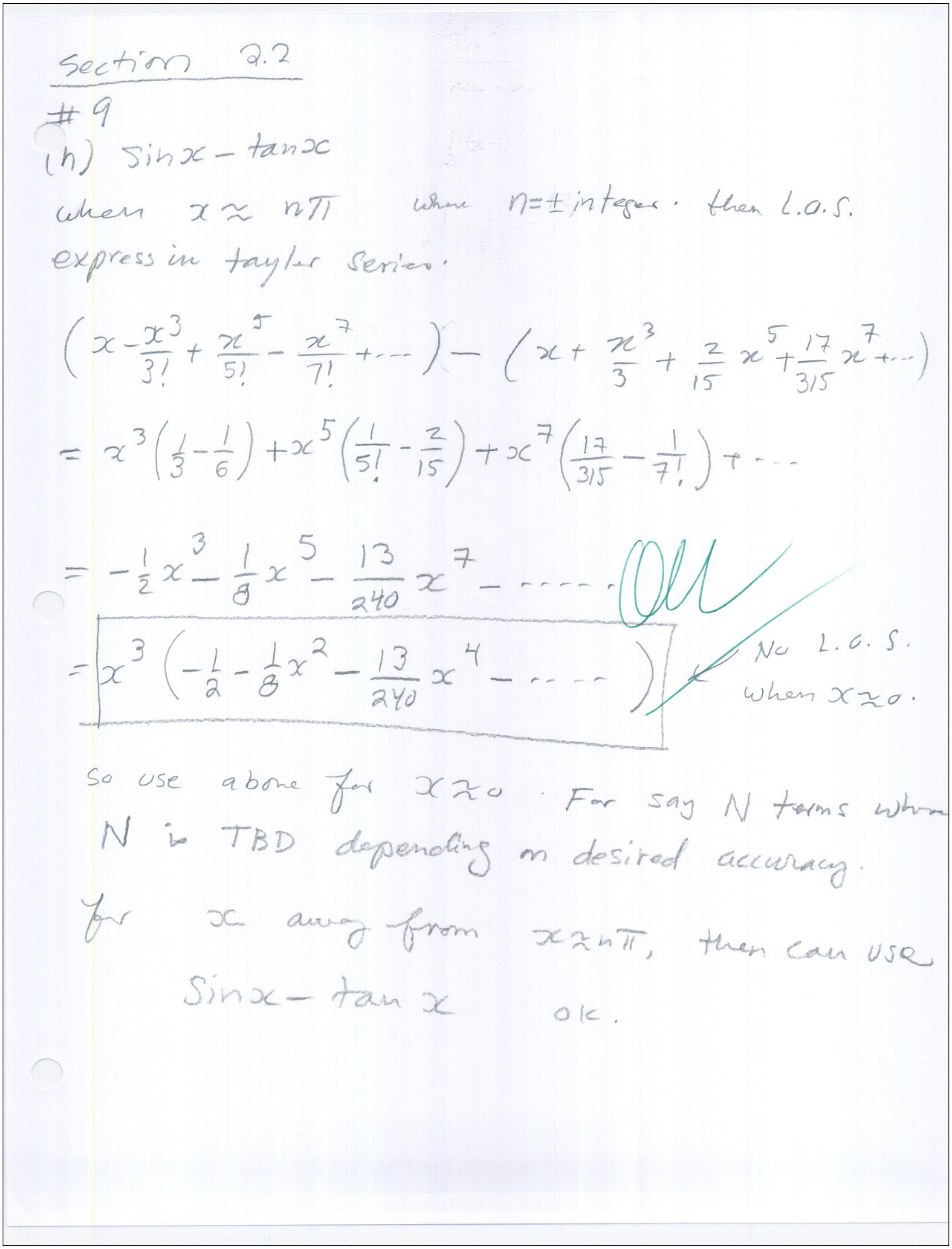
(g) $\frac{\cos x - e^{-x}}{\sin x}$ when $x \sim 0$ loss of significance.

$$= \frac{\left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\right) - \left(1 + (-x) + \frac{(-x)^2}{2!} + \frac{(-x)^3}{3!} + \cdots\right)}{\sin x}$$

$$= \frac{\left(1 - \cancel{\frac{x^2}{2}} + \cancel{\frac{x^4}{4!}} - \frac{x^6}{6!} + \cdots\right) - \left(1 - x + \cancel{\frac{x^2}{2!}} - \frac{x^3}{3!} + \frac{x^4}{4!} - \cdots\right)}{\sin x}$$

$$= \frac{\left(1 - \frac{x^2}{2} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots\right) - 1 + x - \frac{x^2}{2!} + \frac{x^3}{3!} - \frac{x^4}{4!} + \cdots}{\sin x}$$

$$= \frac{x - x^2 + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} - \frac{2}{6!}x^6 \cdots}{\sin x} = \frac{x\left(1 - x + \frac{x^2}{3!} + \frac{x^3}{4!} + \cdots\right)}{\sin x}$$



Section 2.2

#9

(h) $\sin x - \tan x$

when $x \approx n\pi$ where $n = \pm$ integer. then L.O.S.

express in taylor series.

$$\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\cdots\right)-\left(x+\frac{x^3}{3}+\frac{2}{15}x^5+\frac{17}{315}x^7+\cdots\right)$$

$$=x^3\left(\frac{1}{3}-\frac{1}{6}\right)+x^5\left(\frac{1}{5!}-\frac{2}{15}\right)+x^7\left(\frac{17}{315}-\frac{1}{7!}\right)+\cdots$$

$$=-\frac{1}{2}x^3-\frac{1}{8}x^5-\frac{13}{240}x^7-\cdots$$

$$=\boxed{x^3\left(-\frac{1}{2}-\frac{1}{8}x^2-\frac{13}{240}x^4-\cdots\right)}$$

No L.O.S. when $x \approx 0$.

So use above for $x \approx 0$. For say N terms where N is TBD depending on desired accuracy.

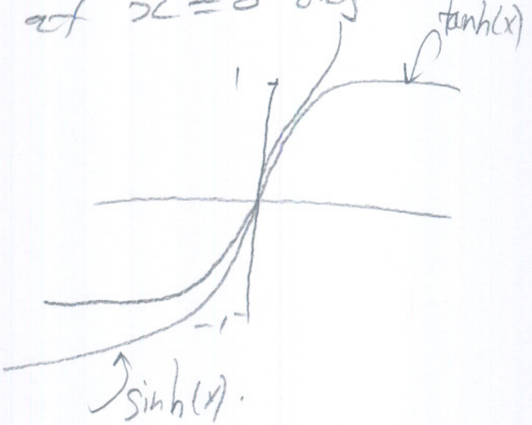
for x away from $x \approx n\pi$, then can use $\sin x - \tan x$ ok.

Section 2.2

9

(i) sinh(x) − tanh(x)

a plot of sinh(x) and tanh(x) shows they are close to each others at x = 0 org.



Taylor: sinh(x) =

$$x+\frac{x^3}{3!}+\frac{x^5}{5!}+\frac{x^7}{7!}+\cdots$$

$$\tanh(x)= x-\frac{x^3}{3}+\frac{2}{15}x^5-\frac{17}{315}x^7+\cdots$$

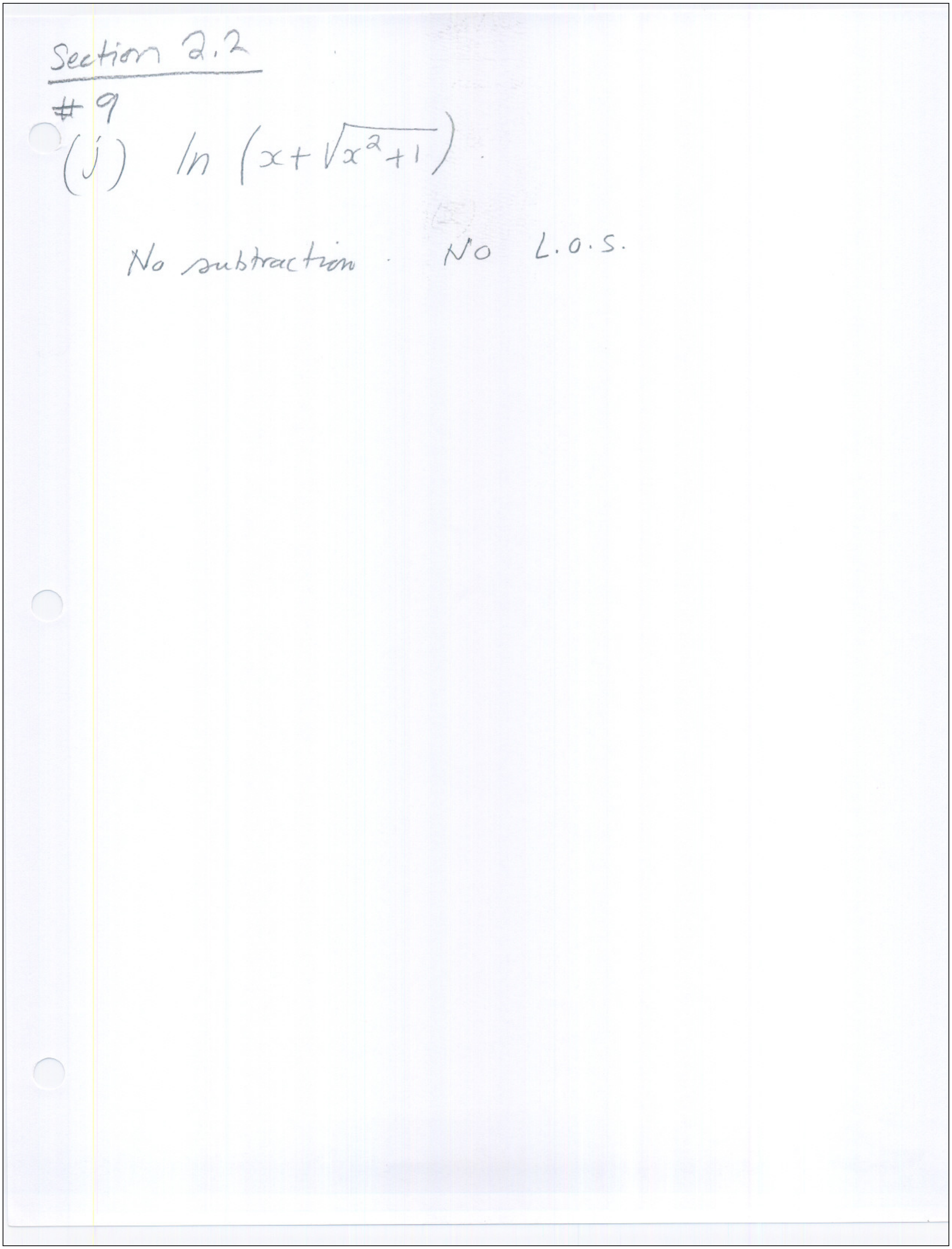
hence $\left(x+\frac{x^3}{3!}+\frac{x^5}{5!}+\frac{x^7}{7!}+\cdots\right)-\left(x-\frac{x^3}{3}+\frac{2}{15}x^5-\frac{17}{315}x^7+\cdots\right)$

$$=x^3\left(\frac{1}{3!}+\frac{1}{3}\right)+x^5\left(\frac{1}{5!}-\frac{2}{15}\right)+x^7\left(\frac{1}{7!}+\frac{17}{315}\right)+\cdots$$

$$=x^3\left(\frac{1}{2}\right)+x^5\left(-\frac{1}{8}\right)+x^7\left(\frac{13}{240}\right)+x^9\left(-\frac{529}{24192}\right)+\cdots$$

$$=x^3\left[\frac{1}{2}-\frac{1}{8}x^2+\frac{13}{240}x^4-\cdots\right]$$

when $x\approx 0$ use ↑ else use sinh(x) − tanh(x).

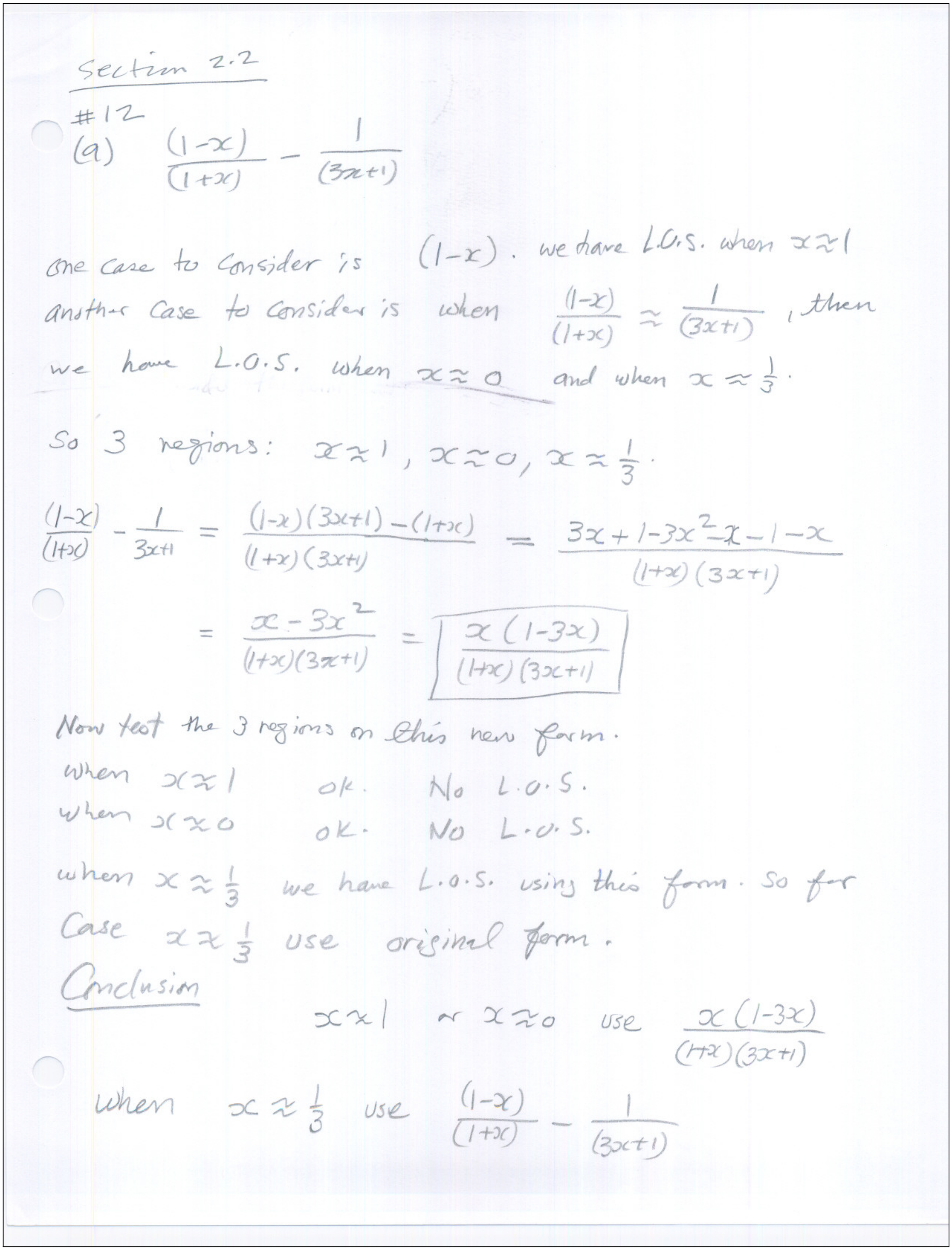


Section 2.2

#9

(j) $\ln\left(x+\sqrt{x^2+1}\right)$.

No subtraction. NO L.O.S.



Section 2.2

#12

(a) $\frac{(1-x)}{(1+x)} - \frac{1}{(3x+1)}$

One case to consider is $(1-x)$. we have L.O.S. when $x \approx 1$

another case to consider is when $\frac{(1-x)}{(1+x)} \approx \frac{1}{(3x+1)}$, then we have L.O.S. when $x \approx 0$ and when $x \approx \frac{1}{3}$.

So 3 regions: $x \approx 1$, $x \approx 0$, $x \approx \frac{1}{3}$.

$$\frac{(1-x)}{(1+x)} - \frac{1}{3x+1} = \frac{(1-x)(3x+1)-(1+x)}{(1+x)(3x+1)} = \frac{3x+1-3x^2-x-1-x}{(1+x)(3x+1)}$$

$$= \frac{x-3x^2}{(1+x)(3x+1)} = \boxed{\frac{x(1-3x)}{(1+x)(3x+1)}}$$

Now test the 3 regions on this new form.

when $x \approx 1$ ok. No L.O.S.

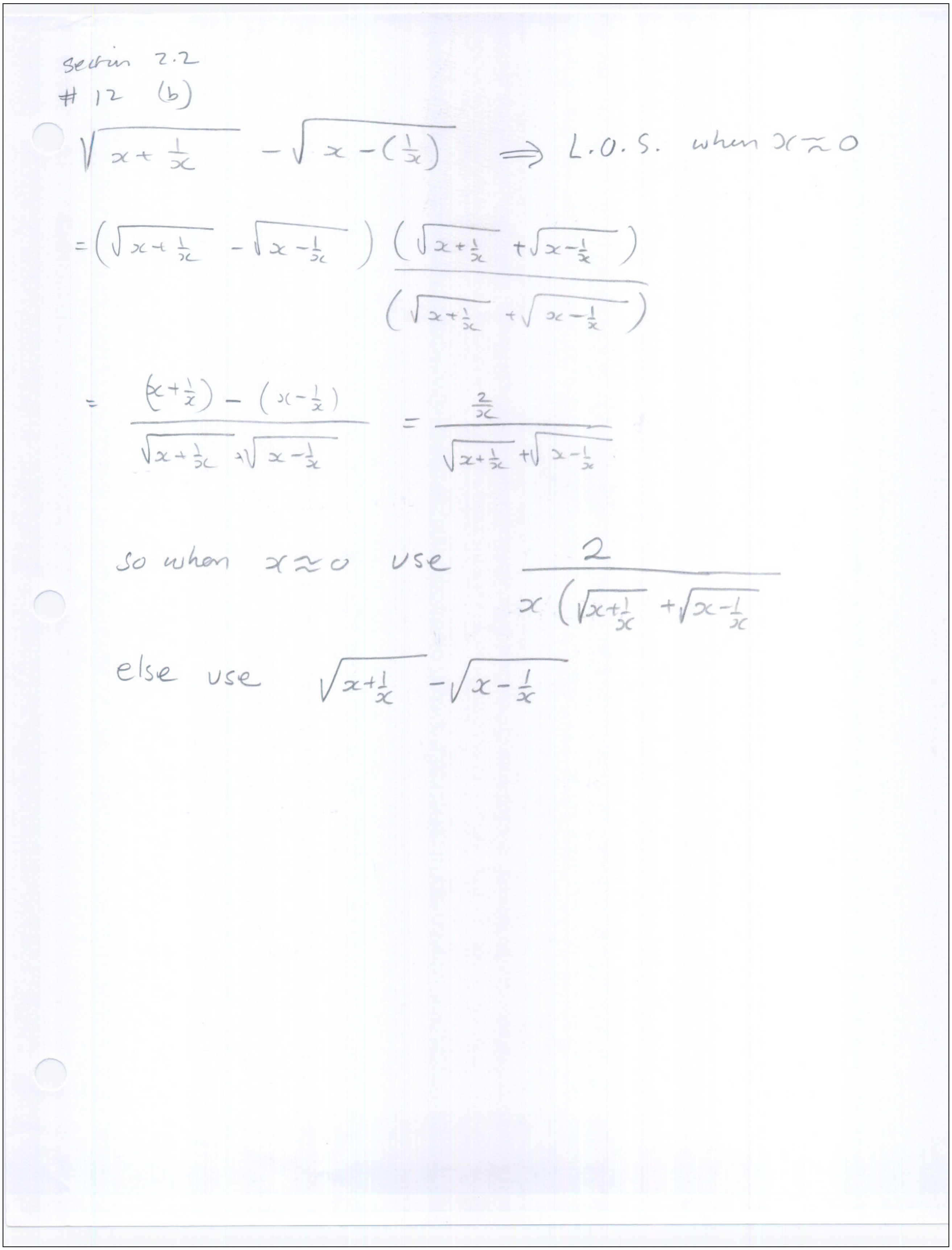
when $x \approx 0$ ok. No L.O.S.

when $x \approx \frac{1}{3}$ we have L.O.S. using this form. So for Case $x \approx \frac{1}{3}$ use original form.

Conclusion

$x \approx 1$ or $x \approx 0$ use $\frac{x(1-3x)}{(1+x)(3x+1)}$

when $x \approx \frac{1}{3}$ use $\frac{(1-x)}{(1+x)} - \frac{1}{(3x+1)}$



section 2.2

#12 (b)

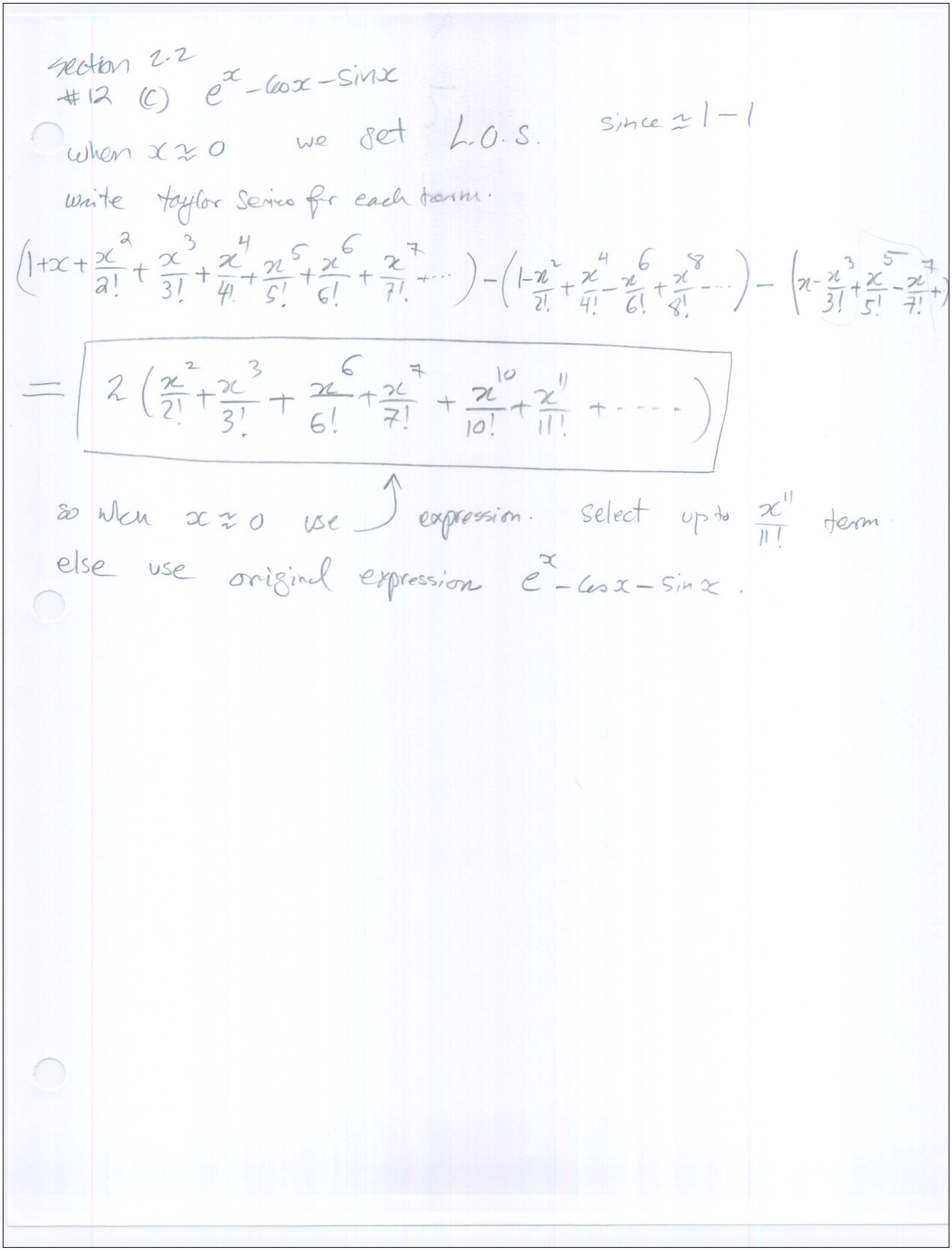
$\sqrt{x+\frac{1}{x}} - \sqrt{x-(\frac{1}{x})} \Rightarrow$ L.O.S. when $x \simeq 0$

$$=\left(\sqrt{x+\frac{1}{x}}-\sqrt{x-\frac{1}{x}}\right)\frac{\left(\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}\right)}{\left(\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}\right)}$$

$$=\frac{(x+\frac{1}{x})-(x-\frac{1}{x})}{\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}}=\frac{\frac{2}{x}}{\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}}$$

so when $x \simeq 0$ use $\dfrac{2}{x\left(\sqrt{x+\frac{1}{x}}+\sqrt{x-\frac{1}{x}}\right.}$

else use $\sqrt{x+\frac{1}{x}}-\sqrt{x-\frac{1}{x}}$



section 2.2

#12 (c) $e^x - \cos x - \sin x$

when $x \approx 0$ we get L.O.S. since $\approx 1-1$

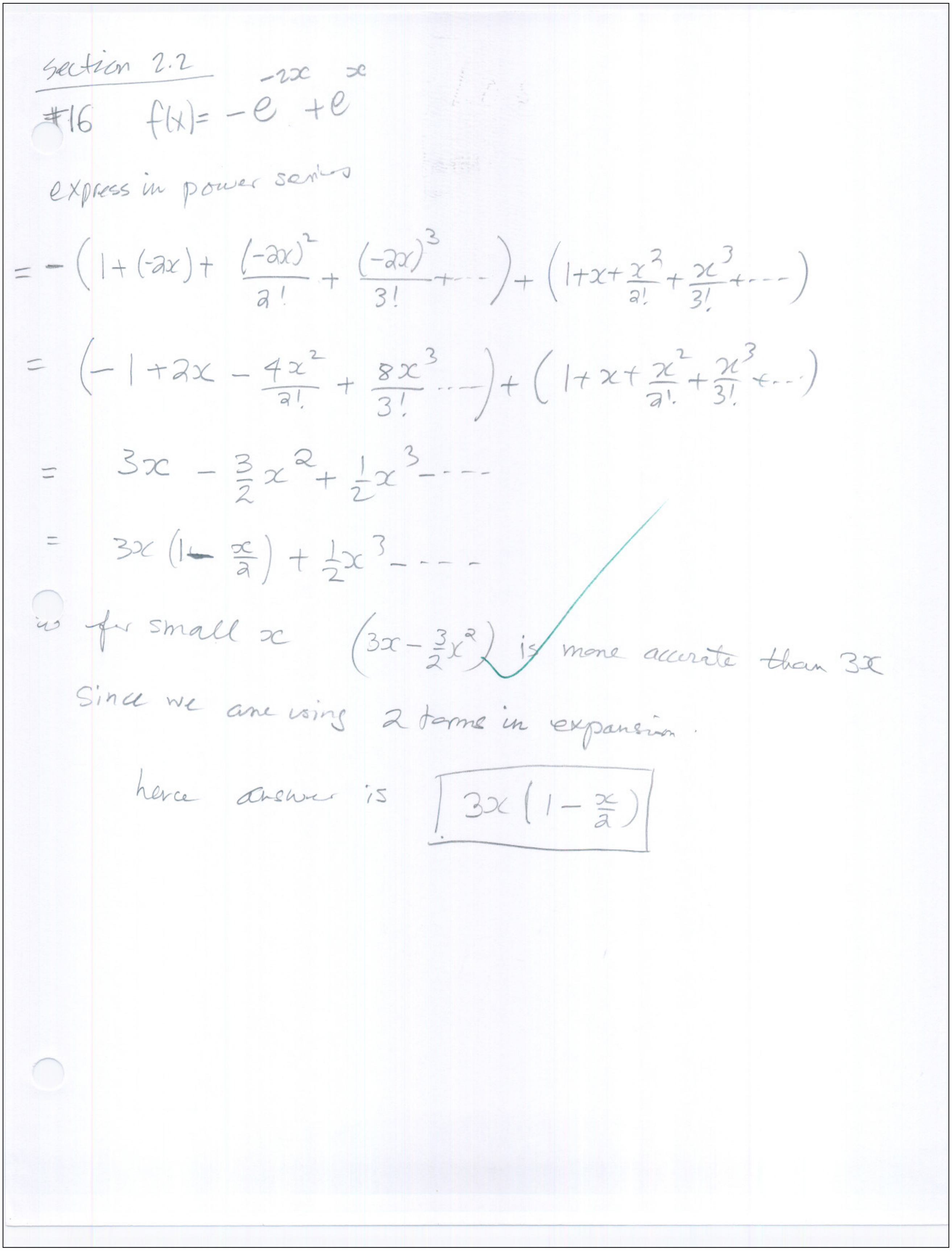
write taylor Series for each term.

$$\left(1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\frac{x^5}{5!}+\frac{x^6}{6!}+\frac{x^7}{7!}+\cdots\right)-\left(1-\frac{x^2}{2!}+\frac{x^4}{4!}-\frac{x^6}{6!}+\frac{x^8}{8!}-\cdots\right)-\left(x-\frac{x^3}{3!}+\frac{x^5}{5!}-\frac{x^7}{7!}+\right)$$

$$=\boxed{2\left(\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^6}{6!}+\frac{x^7}{7!}+\frac{x^{10}}{10!}+\frac{x^{11}}{11!}+\cdots\right)}$$

so when $x \approx 0$ use ↑ expression. Select up to $\frac{x^{11}}{11!}$ term

else use original expression $e^x - \cos x - \sin x$.



section 2.2

#16 $f(x) = -e^{-2x} + e^{x}$

express in power series

$$= -\left(1 + (-2x) + \frac{(-2x)^2}{2!} + \frac{(-2x)^3}{3!} + \cdots\right) + \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)$$

$$= \left(-1 + 2x - \frac{4x^2}{2!} + \frac{8x^3}{3!} \cdots\right) + \left(1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots\right)$$

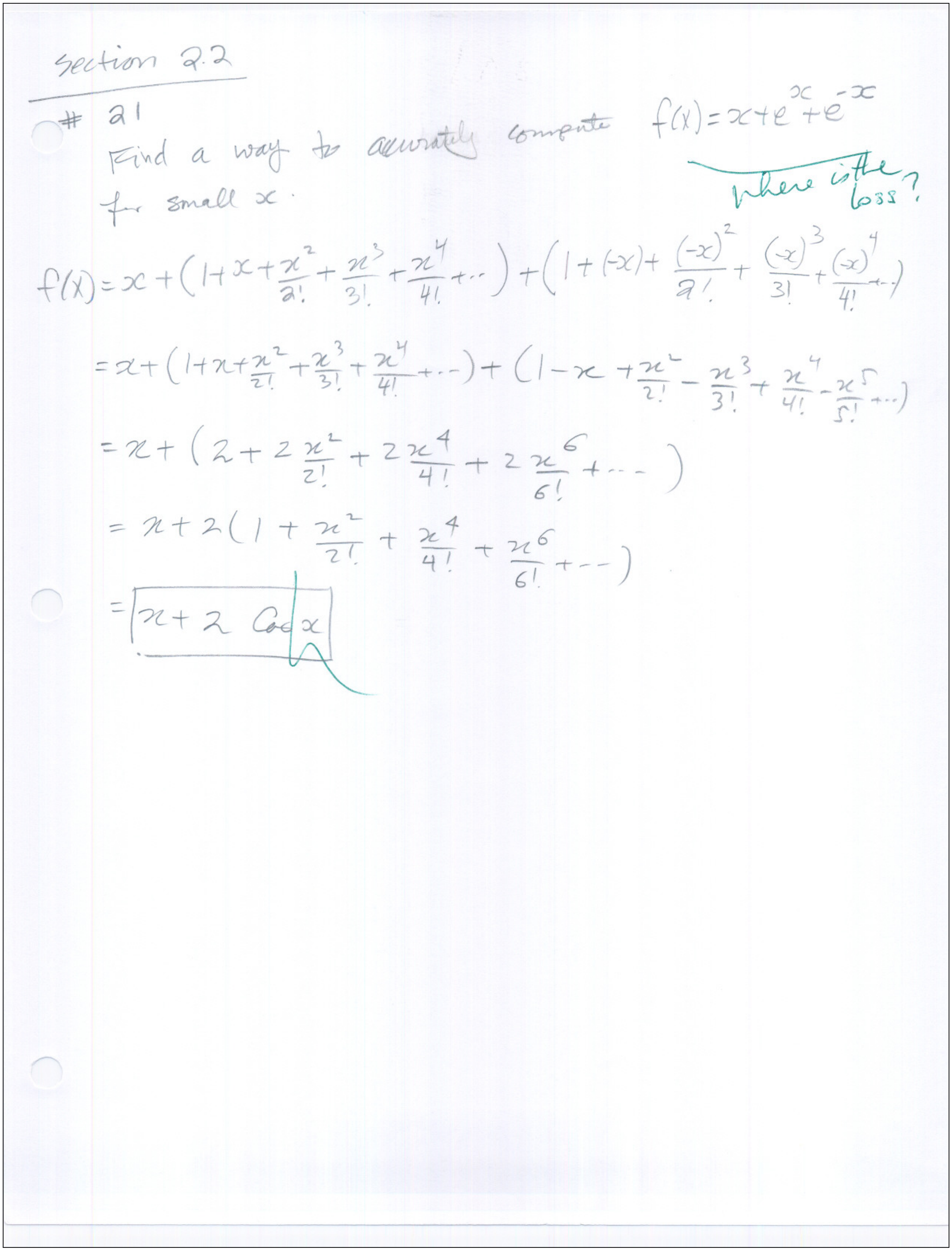
$$= 3x - \frac{3}{2}x^2 + \frac{1}{2}x^3 - \cdots$$

$$= 3x\left(1 - \frac{x}{2}\right) + \frac{1}{2}x^3 - \cdots$$

so for small x $\left(3x - \frac{3}{2}x^2\right)$ is more accurate than $3x$ ✓

Since we are using 2 terms in expansion.

hence answer is $\boxed{3x\left(1 - \frac{x}{2}\right)}$



section 2.2

21

Find a way to accurately compute $f(x)=x+e^{x}+e^{-x}$ for small x.

where is the loss?

$$f(x)=x+\left(1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots\right)+\left(1+(-x)+\frac{(-x)^2}{2!}+\frac{(-x)^3}{3!}+\frac{(-x)^4}{4!}+\cdots\right)$$

$$=x+\left(1+x+\frac{x^2}{2!}+\frac{x^3}{3!}+\frac{x^4}{4!}+\cdots\right)+\left(1-x+\frac{x^2}{2!}-\frac{x^3}{3!}+\frac{x^4}{4!}-\frac{x^5}{5!}+\cdots\right)$$

$$=x+\left(2+2\frac{x^2}{2!}+2\frac{x^4}{4!}+2\frac{x^6}{6!}+\cdots\right)$$

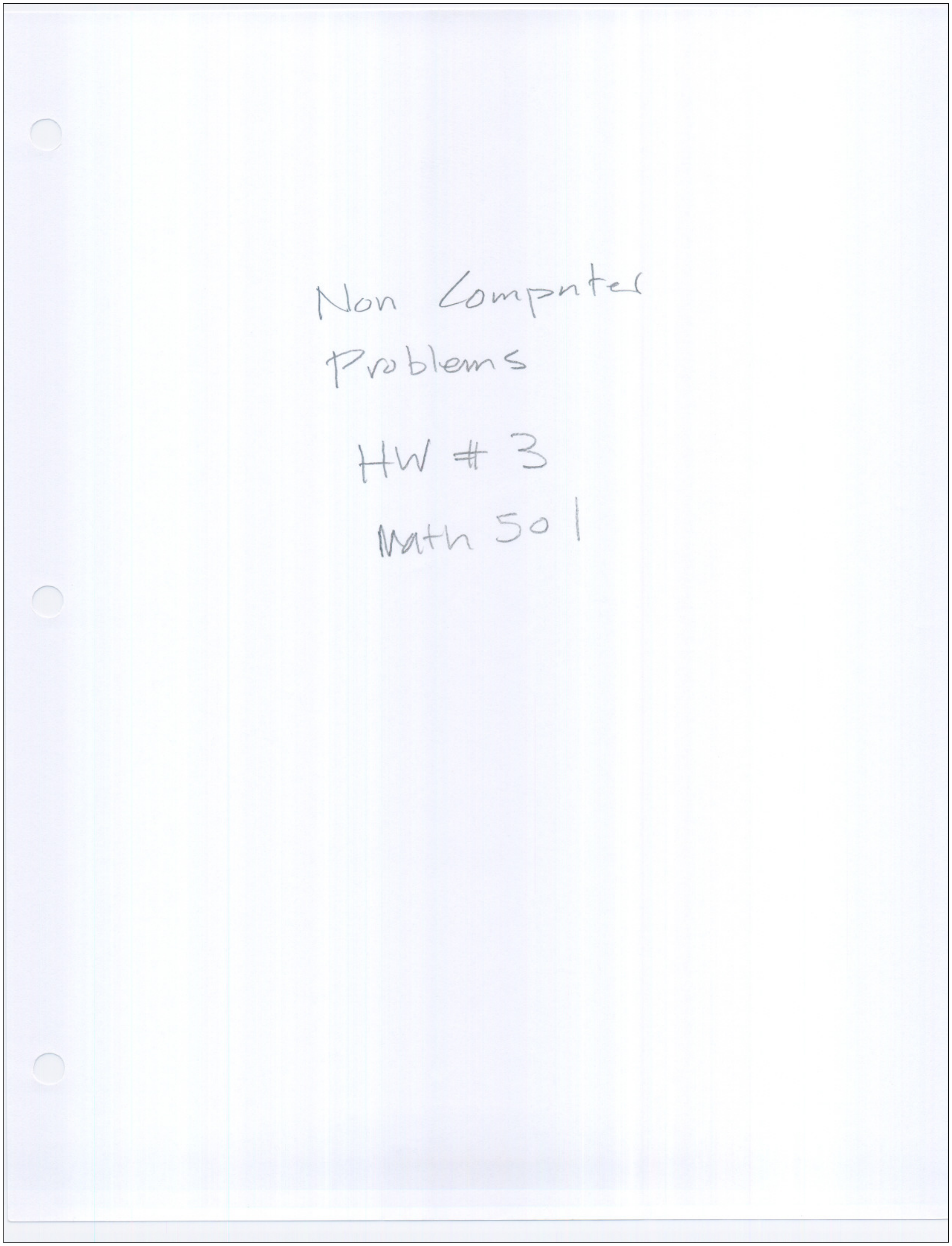
$$=x+2\left(1+\frac{x^2}{2!}+\frac{x^4}{4!}+\frac{x^6}{6!}+\cdots\right)$$

$$=\boxed{x+2\ \mathrm{Cos}\ x}$$

4.4 HW 3

Local contents

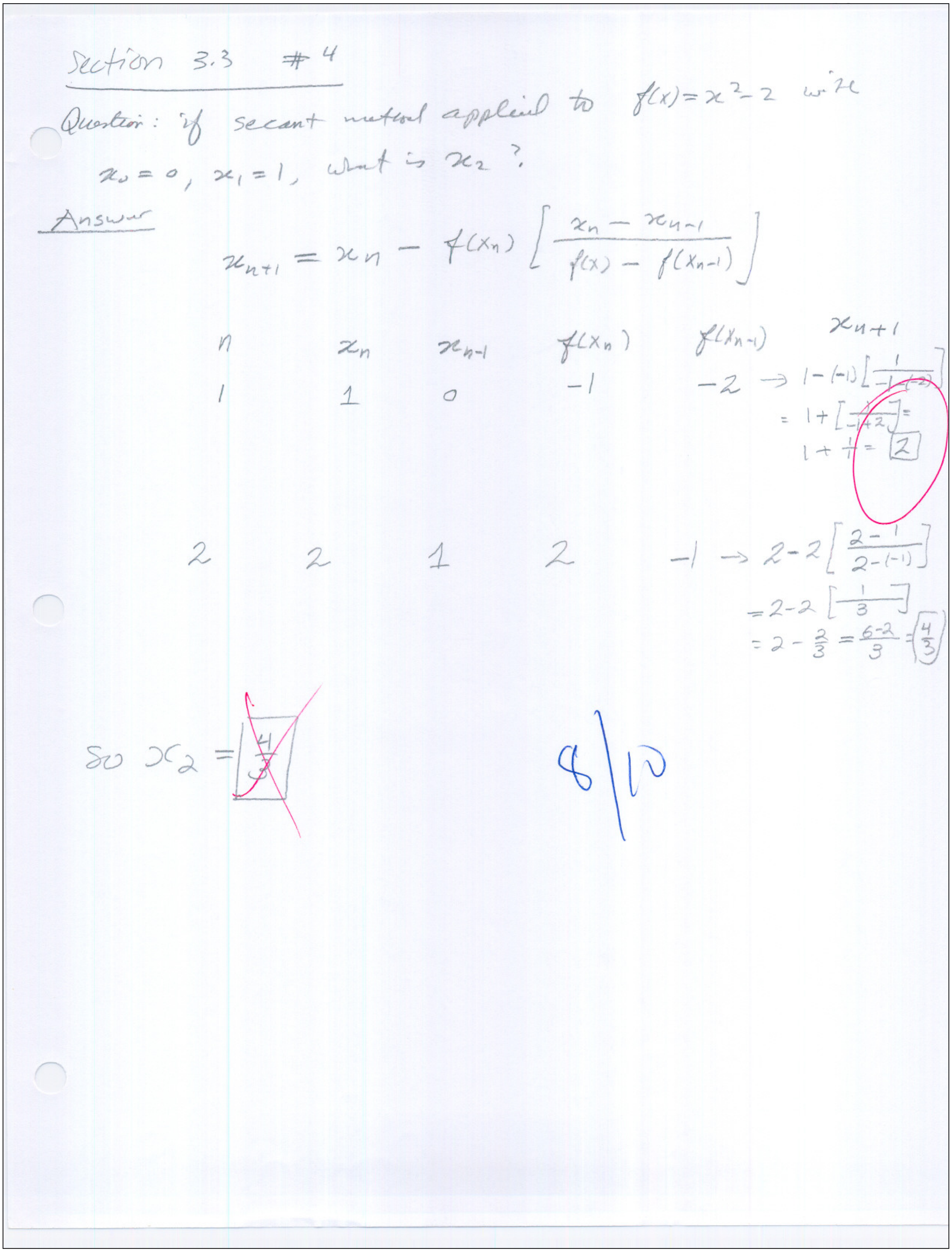
4.4.1 HW 3. analytical part



Non Computer
Problems

HW # 3

Math 501



Section 3.3 # 4

Question: if secant method applied to $f(x)=x^2-2$ with $x_0=0$, $x_1=1$, what is x_2?

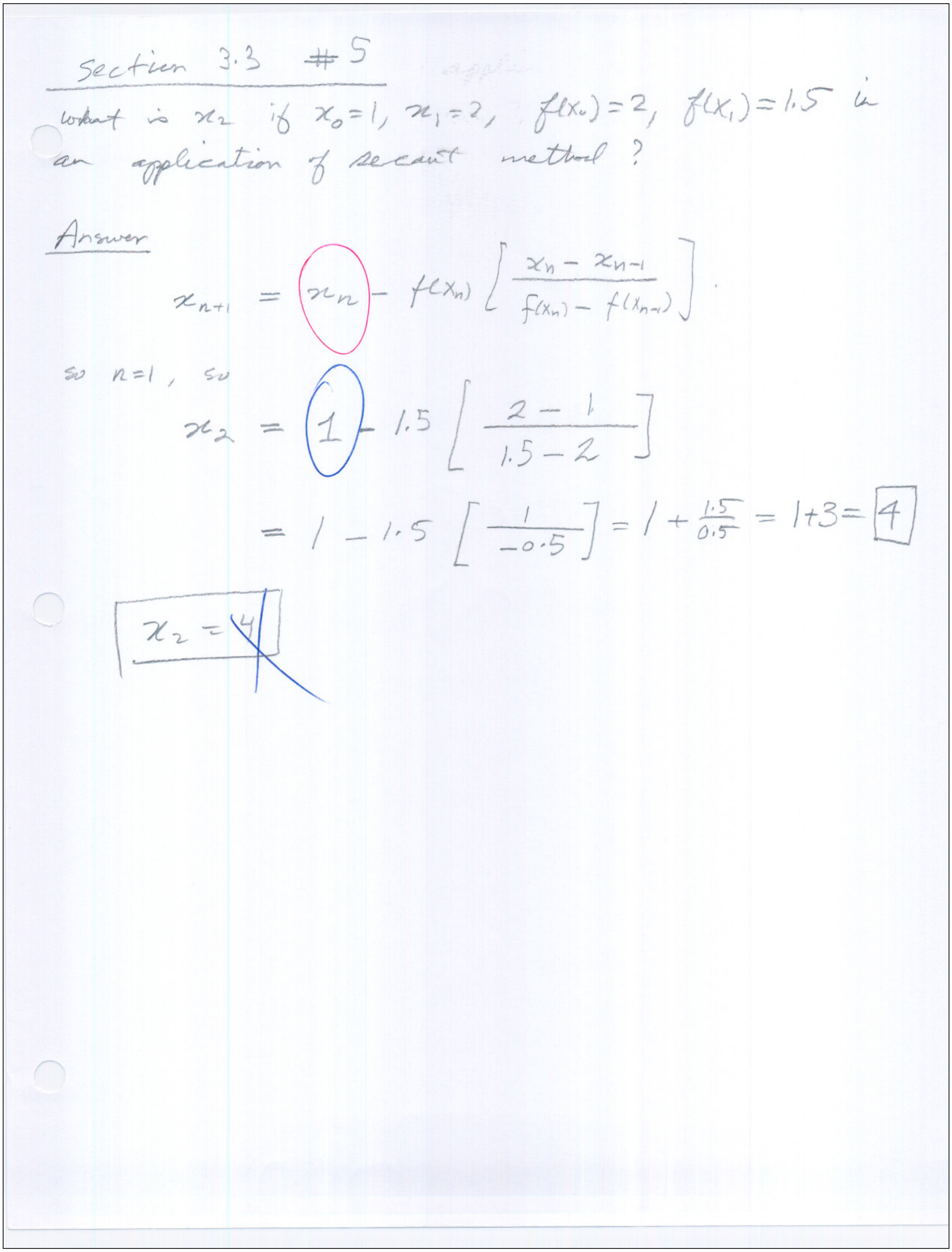
Answer

$$x_{n+1} = x_n - f(x_n)\left[\frac{x_n - x_{n-1}}{f(x) - f(x_{n-1})}\right]$$

n	x_n	x_{n-1}	$f(x_n)$	$f(x_{n-1})$	x_{n+1}
1	1	0	-1	-2	$\rightarrow 1-(-1)\left[\frac{1}{-1-(-2)}\right]$ $= 1+\left[\frac{1}{-1+2}\right] =$ $1+\frac{1}{1} = \boxed{2}$
2	2	1	2	-1	$\rightarrow 2-2\left[\frac{2-1}{2-(-1)}\right]$ $= 2-2\left[\frac{1}{3}\right]$ $= 2-\frac{2}{3} = \frac{6-2}{3} = \boxed{\frac{4}{3}}$

So $x_2 = \cancel{\boxed{\frac{4}{3}}}$

8/10



Section 3.3 #5

what is x_2 if $x_0=1$, $x_1=2$, $f(x_0)=2$, $f(x_1)=1.5$ in an application of secant method ?

Answer

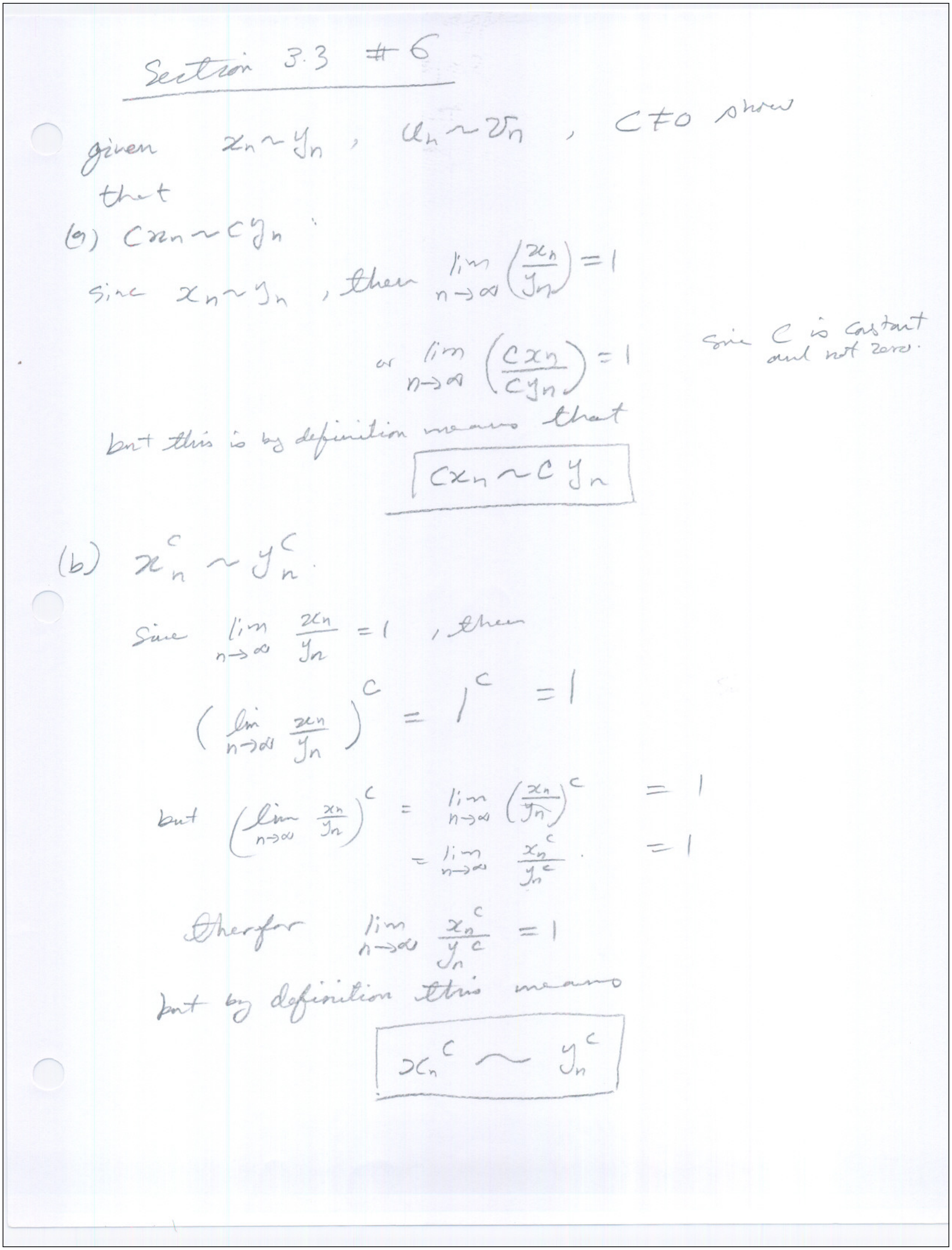
$$x_{n+1} = x_n - f(x_n)\left[\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}\right].$$

so $n=1$, so

$$x_2 = 1 - 1.5\left[\frac{2-1}{1.5-2}\right]$$

$$= 1 - 1.5\left[\frac{1}{-0.5}\right] = 1 + \frac{1.5}{0.5} = 1+3 = \boxed{4}$$

$$\boxed{x_2 = \cancel{4}}$$



Section 3.3 #6

given $x_n \sim y_n$, $u_n \sim v_n$, $c \neq 0$ show that

(a) $cx_n \sim cy_n$:

since $x_n \sim y_n$, then $\lim_{n\to\infty}\left(\frac{x_n}{y_n}\right)=1$

or $\lim_{n\to\infty}\left(\frac{cx_n}{cy_n}\right)=1$ since c is constant and not zero.

but this is by definition means that

$$\boxed{cx_n \sim c\,y_n}$$

(b) $x_n^c \sim y_n^c$.

Since $\lim_{n\to\infty}\frac{x_n}{y_n}=1$, then

$$\left(\lim_{n\to\infty}\frac{x_n}{y_n}\right)^c = 1^c = 1$$

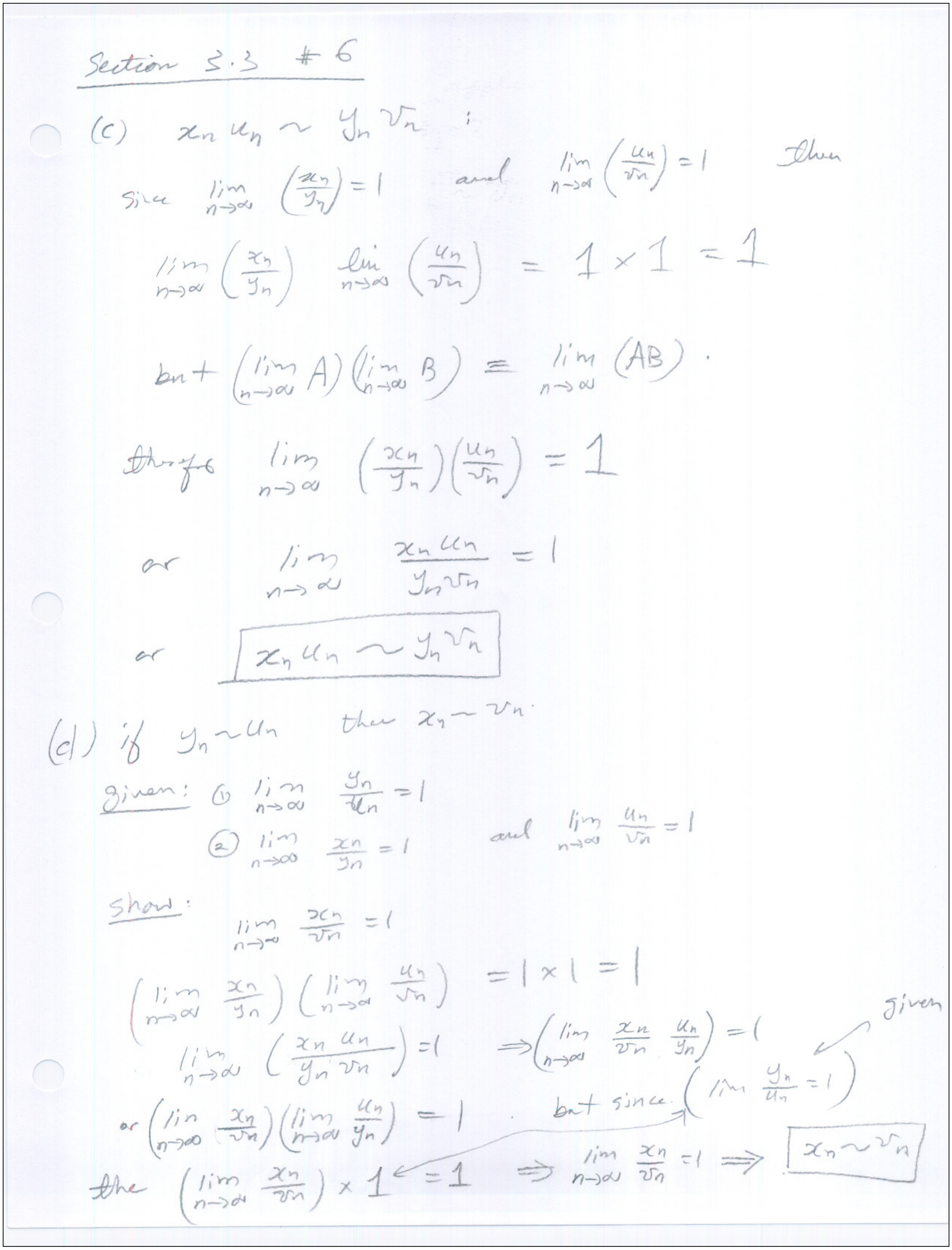
but $\left(\lim_{n\to\infty}\frac{x_n}{y_n}\right)^c = \lim_{n\to\infty}\left(\frac{x_n}{y_n}\right)^c = 1$

$= \lim_{n\to\infty}\frac{x_n^c}{y_n^c}. = 1$

Therfor $\lim_{n\to\infty}\frac{x_n^c}{y_n^c} = 1$

but by definition this means

$$\boxed{x_n^c \sim y_n^c}$$



Section 3.3 # 6

(c) $x_n u_n \sim y_n v_n$:

Since $\lim\limits_{n\to\infty}\left(\frac{x_n}{y_n}\right)=1$ and $\lim\limits_{n\to\infty}\left(\frac{u_n}{v_n}\right)=1$ then

$$\lim_{n\to\infty}\left(\frac{x_n}{y_n}\right)\lim_{n\to\infty}\left(\frac{u_n}{v_n}\right)=1\times 1=1$$

but $\left(\lim\limits_{n\to\infty} A\right)\left(\lim\limits_{n\to\infty} B\right)=\lim\limits_{n\to\infty}(AB)$.

therefore $\lim\limits_{n\to\infty}\left(\frac{x_n}{y_n}\right)\left(\frac{u_n}{v_n}\right)=1$

or $\lim\limits_{n\to\infty}\frac{x_n u_n}{y_n v_n}=1$

or $\boxed{x_n u_n \sim y_n v_n}$

(d) if $y_n \sim u_n$ then $x_n \sim v_n$.

given: ① $\lim\limits_{n\to\infty}\frac{y_n}{u_n}=1$

② $\lim\limits_{n\to\infty}\frac{x_n}{y_n}=1$ and $\lim\limits_{n\to\infty}\frac{u_n}{v_n}=1$

show: $\lim\limits_{n\to\infty}\frac{x_n}{v_n}=1$

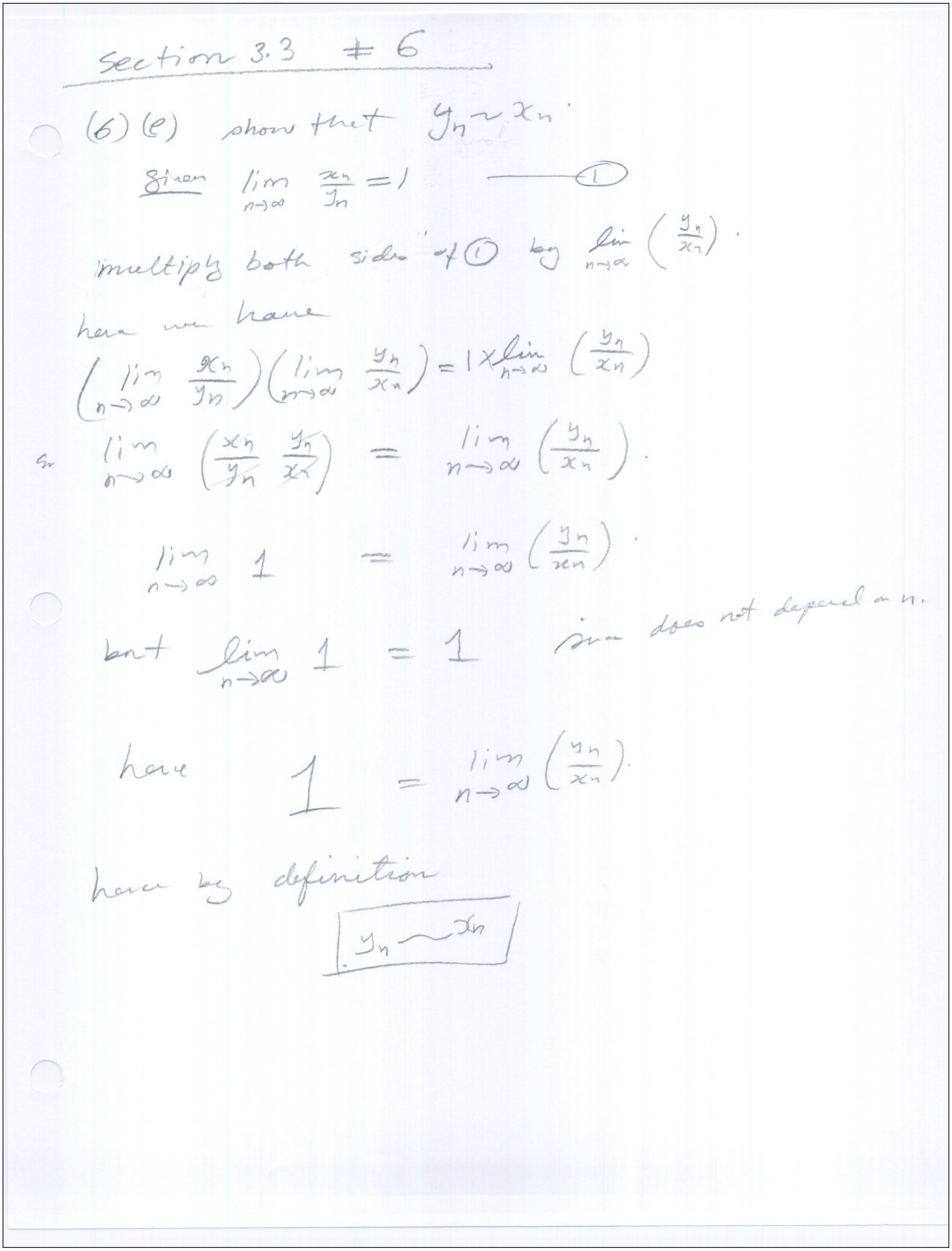
$$\left(\lim_{n\to\infty}\frac{x_n}{y_n}\right)\left(\lim_{n\to\infty}\frac{u_n}{v_n}\right)=1\times 1=1$$

$$\lim_{n\to\infty}\left(\frac{x_n u_n}{y_n v_n}\right)=1 \Rightarrow \left(\lim_{n\to\infty}\frac{x_n}{v_n}\frac{u_n}{y_n}\right)=1$$

given

or $\left(\lim\limits_{n\to\infty}\frac{x_n}{v_n}\right)\left(\lim\limits_{n\to\infty}\frac{u_n}{y_n}\right)=1$ but since $\left(\lim\frac{y_n}{u_n}=1\right)$

the $\left(\lim\limits_{n\to\infty}\frac{x_n}{v_n}\right)\times 1=1 \Rightarrow \lim\limits_{n\to\infty}\frac{x_n}{v_n}=1 \Rightarrow \boxed{x_n\sim v_n}$



section 3.3 # 6

(6) (e) show that $y_n \sim x_n$

given $\lim_{n\to\infty} \frac{x_n}{y_n} = 1$ ——— ①

multiply both sides of ① by $\lim_{n\to\infty}\left(\frac{y_n}{x_n}\right)$.

hence we have

$$\left(\lim_{n\to\infty} \frac{x_n}{y_n}\right)\left(\lim_{n\to\infty} \frac{y_n}{x_n}\right) = 1 \times \lim_{n\to\infty}\left(\frac{y_n}{x_n}\right)$$

So $\lim_{n\to\infty}\left(\frac{\cancel{x_n}}{\cancel{y_n}} \frac{\cancel{y_n}}{\cancel{x_n}}\right) = \lim_{n\to\infty}\left(\frac{y_n}{x_n}\right)$.

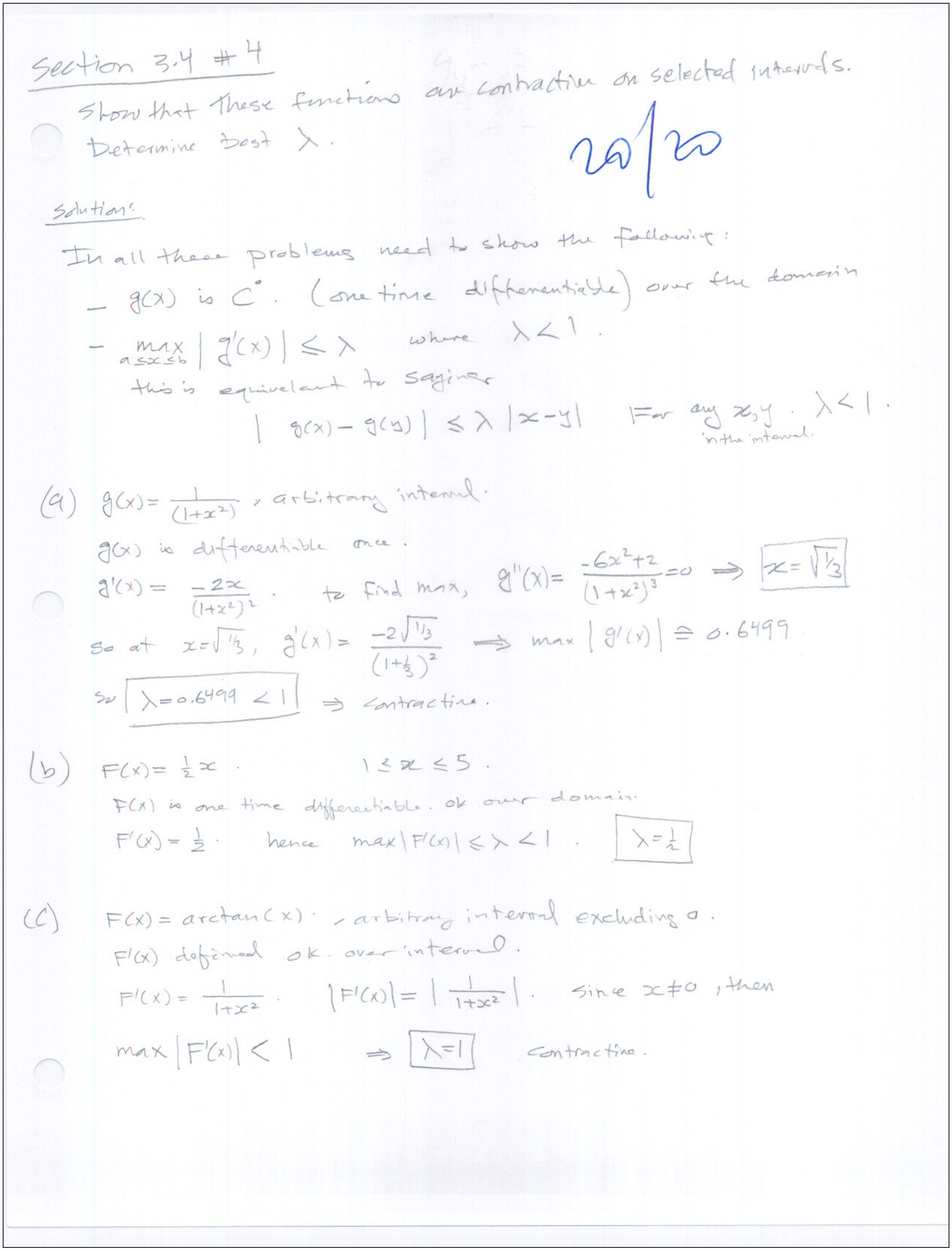
$$\lim_{n\to\infty} 1 = \lim_{n\to\infty}\left(\frac{y_n}{x_n}\right).$$

but $\lim_{n\to\infty} 1 = 1$ since does not depend on n.

hence $1 = \lim_{n\to\infty}\left(\frac{y_n}{x_n}\right)$.

hence by definition

$$\boxed{y_n \sim x_n}$$



Section 3.4 #4

Show that these functions are contraction on selected intervals. Determine best λ.

20/20

Solution:

In all these problems need to show the following:

- $g(x)$ is C^{0}. (one time differentiable) over the domain
- $\max\limits_{a\le x\le b} |g'(x)| \le \lambda$ where $\lambda < 1$.

this is equivalent to saying

$$|g(x)-g(y)| \le \lambda |x-y| \quad \text{For any } x,y \text{ in the interval}. \quad \lambda<1.$$

(a) $g(x)=\frac{1}{(1+x^2)}$, arbitrary interval.

$g(x)$ is differentiable once.

$g'(x) = \frac{-2x}{(1+x^2)^2}$. to find max, $g''(x)=\frac{-6x^2+2}{(1+x^2)^3}=0 \Rightarrow \boxed{x=\sqrt{1/3}}$

So at $x=\sqrt{1/3}$, $g'(x)=\frac{-2\sqrt{1/3}}{(1+\frac{1}{3})^2} \Rightarrow \max|g'(x)| \mathrel{\hat{=}} 0.6499$

So $\boxed{\lambda=0.6499<1} \Rightarrow$ contractive.

(b) $F(x)=\frac{1}{2}x$. $1\le x\le 5$.

$F(x)$ is one time differentiable. ok over domain.

$F'(x)=\frac{1}{2}$. hence $\max|F'(x)|\le\lambda<1$. $\boxed{\lambda=\frac{1}{2}}$

(c) $F(x)=\arctan(x)$. , arbitrary interval excluding 0.

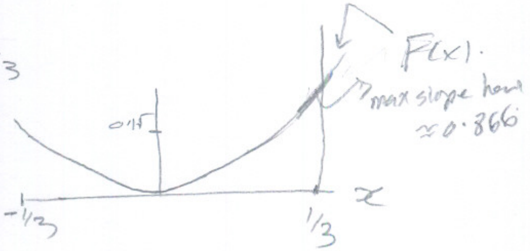
$F'(x)$ defined ok. over interval.

$F'(x)=\frac{1}{1+x^2}$. $|F'(x)|=\left|\frac{1}{1+x^2}\right|$. since $x\neq 0$, then

$\max|F'(x)|<1 \Rightarrow \boxed{\lambda=1}$ Contractive.

section 3.4 # 4

(d) $F(x) = |x|^{3/2}$ on $|x| \leq 1/3$



$F'(x)$ continuous on domain ok.

Consider positive range.

$F'(x) = \frac{3}{2} x^{1/2}$. This is max when x is max. i.e $x = 1/3$.

so max $F(x) = \frac{3}{2}(1/3)^{1/2} \equiv 0.866$--

by symmetry of $F(x) \Rightarrow \underset{|x| \leq 1/3}{\underset{max}{|F'(x)|}} \leq \lambda < 1$ where $\boxed{\lambda \hat{=} 0.866-}$

$\rightarrow$ contractive.

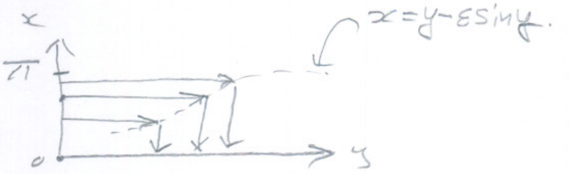
section 3.4 #5

Kepler equation $x = y - \varepsilon \sin y$ $0 < \varepsilon < 1$.

show that for each $x \in [0, \pi]$ there is a y satisfying the equation.

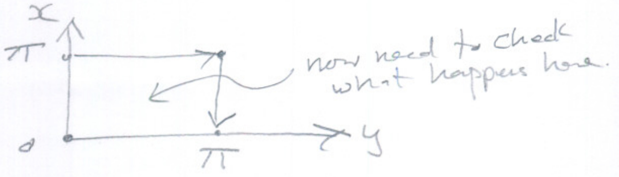
answer

we need to show that if we pick any $x \in [0,\pi]$ value, then we can map that value to some y value via function $y - \varepsilon \sin y$.

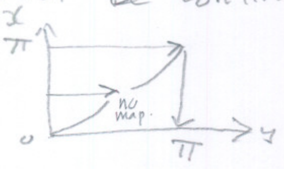


First, when $x = 0$, then $y = 0$ satisfy the equation.

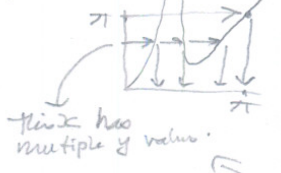
at $x = \pi$, then $y = \pi$ satisfy the equation.



now between $0, \pi$, for any x to map to a y, then $y - \varepsilon \sin y$ must be continous over $[0, \pi]$ to avoid case such as



In addition to avoid case that x maps to more than one y value, need to avoid case such as



we see that $y - \varepsilon \sin y$ is continouse since its derivative is $1 - \varepsilon \cos y$ which defined over all $[0, \pi]$. to handle this second case need to show $y - \varepsilon \sin y$ has +ve derivative and < 1:

$g(y) = y - \varepsilon \sin y \Rightarrow g'(y) = 1 - \varepsilon \cos y \Rightarrow g'(y) = 1 - \varepsilon \cos y$. but $|\cos y| \leq 1$

so $\boxed{0 < g'(y) < 1}$ since $|\varepsilon| < 1$ hence $g(y)$ has +ve slope < 1, hence satisfy second case. QED

Section 3.4 #10

if we attempt to find a fixed point of F by using Newton's method on the equation $F(x)-x=0$, what iteration formula results?

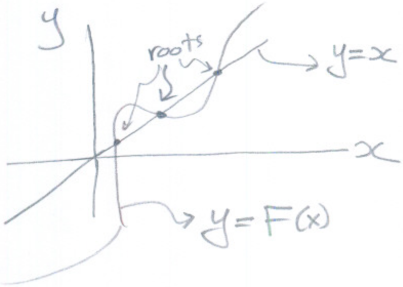
Answer

Newton iterative formula is

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

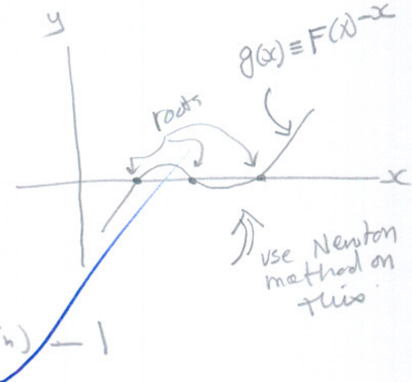
in this problem we seek to find root for $F(x)-x=0$

hence $g(x) = F(x) - x$ will cross the x-axis at the place where $y=F(x)$ and $y=x$ intersect.



So when using Newton method, replace $f(x)$ in that formula by $g(x)$. This results in:

$$x_{n+1} = x_n - \frac{g(x)|_{x=x_n}}{g'(x)|_{x=x_n}}$$

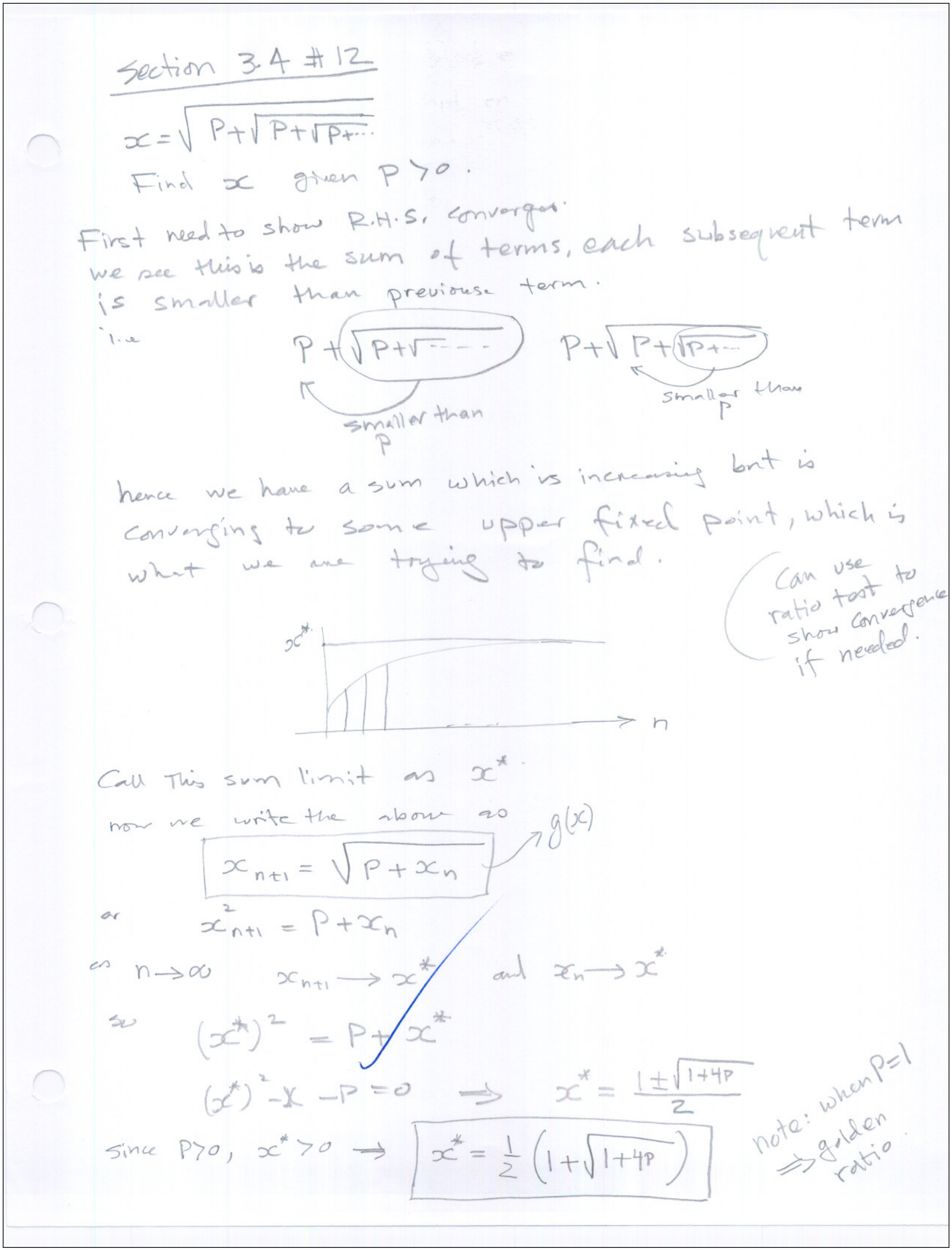


but $g(x)|_{x=x_n} = F(x_n) - x_n$

$$g'(x)|_{x=x_n} = F'(x)|_{x=x_n} - 1 = F'(x_n) - 1$$

so

$$\boxed{x_{n+1} = x_n - \frac{F(x_n)-x_n}{F'(x_n)-1}}$$



Section 3.4 #12

$$x=\sqrt{P+\sqrt{P+\sqrt{P+\cdots}}}$$

Find x given $P>0$.

First need to show R.H.S. converges. we see this is the sum of terms, each subsequent term is smaller than previouse term.

i.e.

$$P+\sqrt{P+\sqrt{\cdots}} \qquad P+\sqrt{P+\sqrt{P+\cdots}}$$

smaller than P

Smaller than P

hence we have a sum which is increasing but is converging to some upper fixed point, which is what we are trying to find.

(Can use ratio test to show convergence if needed.)

Call This sum limit as x^*.

now we write the above as

$$\boxed{x_{n+1}=\sqrt{P+x_n}} \rightarrow g(x)$$

or

$$x_{n+1}^2=P+x_n$$

as $n\to\infty$ $\quad x_{n+1}\to x^*$ and $x_n\to x^*$

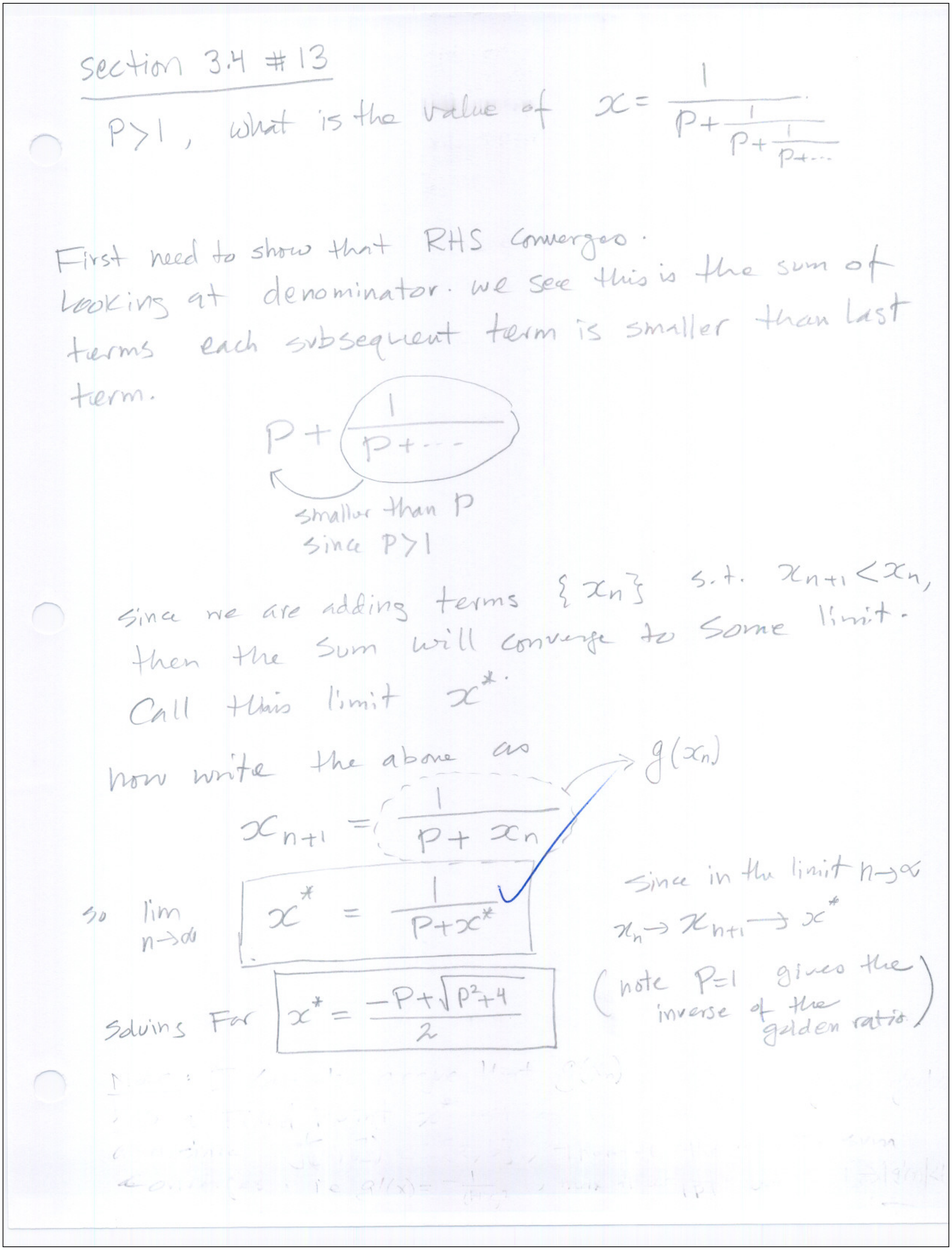
so

$$(x^*)^2=P+x^* \checkmark$$

$$(x^*)^2-x-P=0 \implies x^*=\frac{1\pm\sqrt{1+4P}}{2}$$

since $P>0$, $x^*>0 \implies \boxed{x^*=\frac{1}{2}\left(1+\sqrt{1+4P}\right)}$

note: when $P=1$ $\implies$ golden ratio.



section 3.4 #13

P>1, what is the value of $x = \frac{1}{P+\frac{1}{P+\frac{1}{P+\cdots}}}$

First need to show that RHS converges. Looking at denominator. we see this is the sum of terms each subsequent term is smaller than last term.

$$P + \underbrace{\frac{1}{P+\cdots}}_{\text{smaller than } P \text{ since } P>1}$$

Since we are adding terms $\{x_n\}$ s.t. $x_{n+1} < x_n$, then the sum will converge to some limit. Call this limit x^*.

now write the above as

$$x_{n+1} = \underbrace{\frac{1}{P+x_n}}_{g(x_n)}$$

so $\lim_{n\to\infty}$ $\boxed{x^* = \frac{1}{P+x^*}}$ ✓

since in the limit $n \to \infty$ $x_n \to x_{n+1} \to x^*$

Solving for $\boxed{x^* = \frac{-P+\sqrt{P^2+4}}{2}}$

(note P=1 gives the inverse of the golden ratio)

section 3.4 #29

Prove that $F(x) = 2+x-\arctan(x)$ has property $|F'(x)| < 1$. prove that $F(x)$ does not have a Fixed point.

Answer

$$F'(x) = 1 - \frac{1}{1+x^2} = \frac{x^2}{1+x^2}$$

hence $|F'(x)| < 1$

For a Function $F(x)$ which $|F'(x)| < 1$ not to have a Fixed point, then we need to show that

$F(x) - x = 0$ has No solution.

i.e $\boxed{g(x) = F(x) - x}$ is alway positive or always negative. meaning it never cross the x-axis. hence no root.

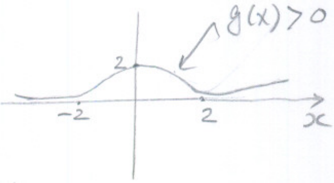
$g(x) = 2 + x - \arctan(x) - x = 2 - \arctan(x)$

Now $\arctan(x)$ will range between $[0, \frac{\pi}{2}]$ or $[0, -\frac{\pi}{2}]$.

but $|\frac{\pi}{2}| \simeq 1.57079\ldots$

hence min $g(x) = 2 - \frac{\pi}{2} = 0.4292\ldots$

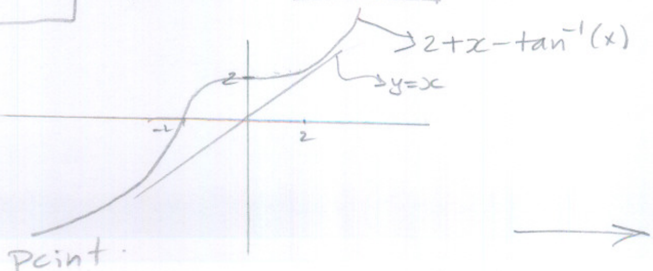
max $g(x) = 2 + \frac{\pi}{2} = 3.5707$.



$\Rightarrow \boxed{g(x) > 0 \quad \text{For all } x}$ ThereFor No root

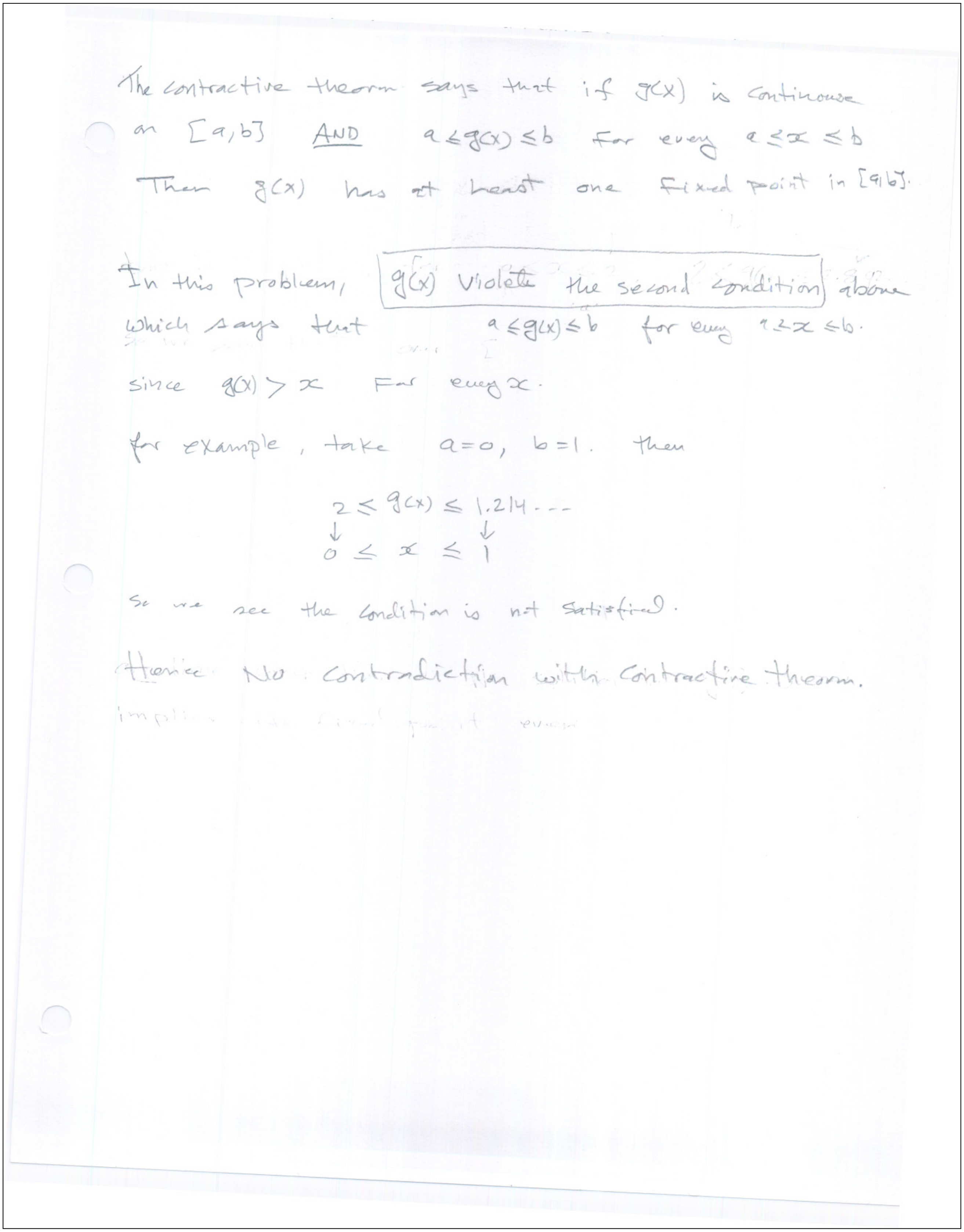
graphically:

The line $y=x$ never intersects $F(x)$.



$\Rightarrow$ No Fixed point.

→



The contractive theorm says that if $g(x)$ is continouse on $[a,b]$ AND $a \le g(x) \le b$ for every $a \le x \le b$ Then $g(x)$ has at least one fixed point in $[a,b]$.

In this problem, $\boxed{g(x) \text{ violate the second condition}}$ above which says that $a \le g(x) \le b$ for every $a \le x \le b$.

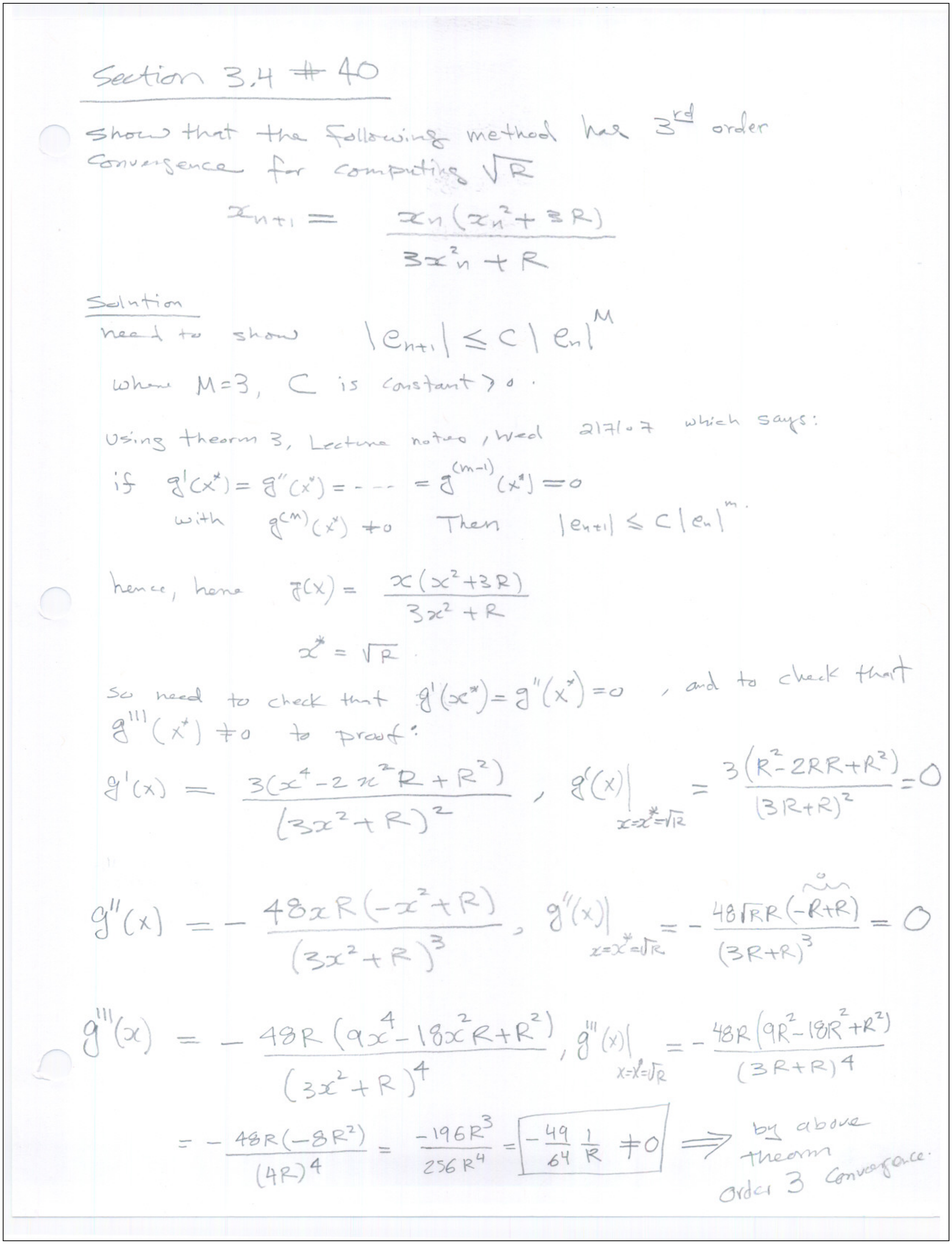
since $g(x) > x$ For every x.

for example, take $a=0$, $b=1$. then

$$\begin{array}{ccccc} 2 & \le & g(x) & \le & 1.214\ldots \\ \downarrow & & & & \downarrow \\ 0 & \le & x & \le & 1 \end{array}$$

So we see the condition is not satisfied.

Hence No contradiction with contractive theorm.



Section 3.4 #40

Show that the following method has 3^{rd} order Convergence for computing $\sqrt{R}$

$$x_{n+1} = \frac{x_n(x_n^2+3R)}{3x_n^2+R}$$

Solution

need to show $|e_{n+1}| \leq C|e_n|^M$

where $M=3$, C is constant >0.

Using theorm 3, Lecture notes, Wed 2/7/07 which says:

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

with $g^{(m)}(x^*) \neq 0$ Then $|e_{n+1}| \leq C|e_n|^m$.

hence, here $g(x) = \frac{x(x^2+3R)}{3x^2+R}$

$x^* = \sqrt{R}$

So need to check that $g'(x^*) = g''(x^*) = 0$, and to check that $g'''(x^*) \neq 0$ to proof:

$$g'(x) = \frac{3(x^4-2x^2R+R^2)}{(3x^2+R)^2}, \quad g'(x)\Big|_{x=x^*=\sqrt{R}} = \frac{3(R^2-2RR+R^2)}{(3R+R)^2} = 0$$

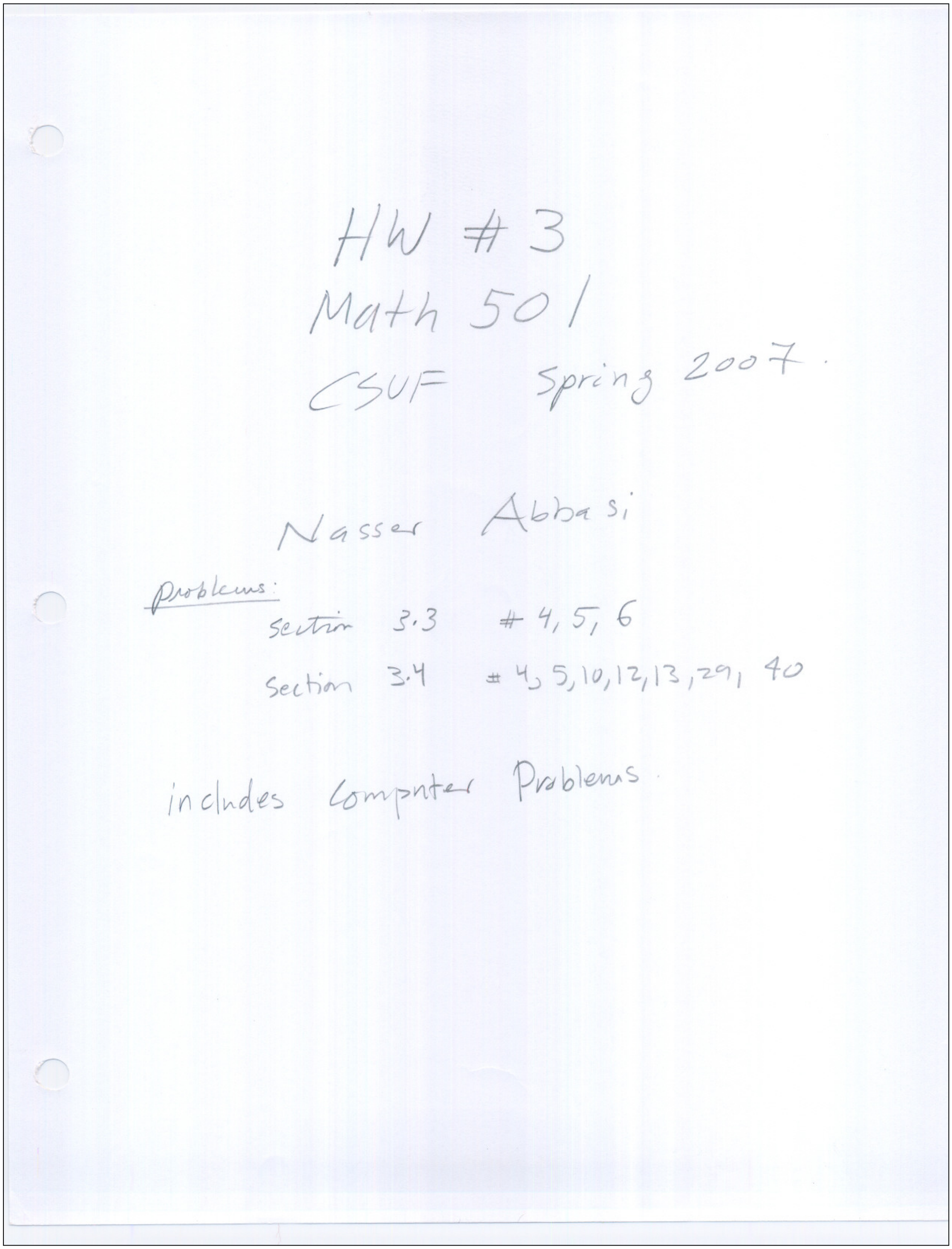
$$g''(x) = -\frac{48xR(-x^2+R)}{(3x^2+R)^3}, \quad g''(x)\Big|_{x=x^*=\sqrt{R}} = -\frac{48\sqrt{R}R(\overbrace{-R+R}^{0})}{(3R+R)^3} = 0$$

$$g'''(x) = -\frac{48R(9x^4-18x^2R+R^2)}{(3x^2+R)^4}, \quad g'''(x)\Big|_{x=x^*=\sqrt{R}} = -\frac{48R(9R^2-18R^2+R^2)}{(3R+R)^4}$$

$$= -\frac{48R(-8R^2)}{(4R)^4} = \frac{-196R^3}{256R^4} = \boxed{-\frac{49}{64}\frac{1}{R} \neq 0} \Longrightarrow$$

by above theorm order 3 Convergence.

4.4.2 HW 3. Computer part Matlab horner method, Taylor approx, Bisection and secant



HW #3

Math 501

CSUF Spring 2007.

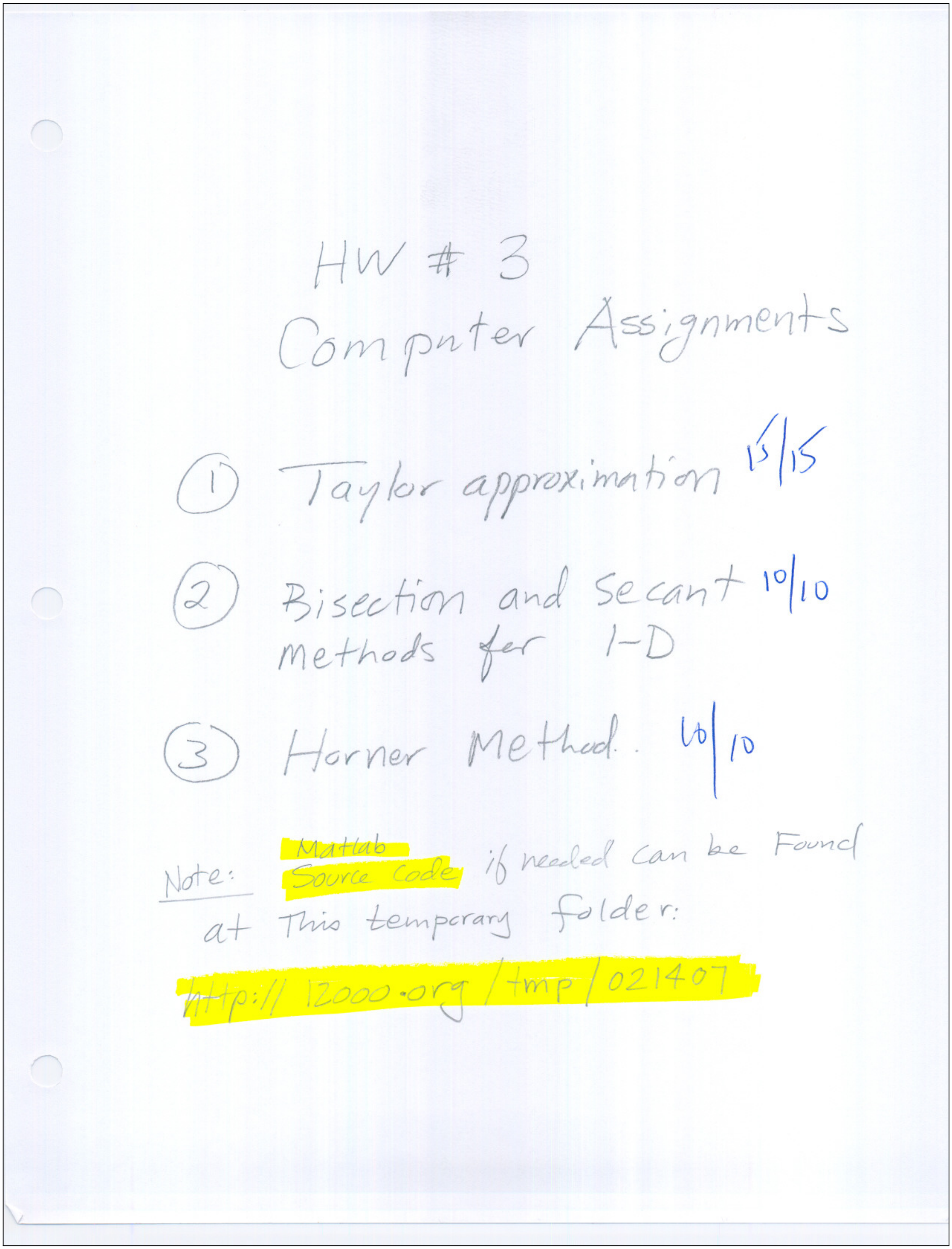
Nasser Abbasi

problems:

section 3.3 # 4, 5, 6

Section 3.4 # 4, 5, 10, 12, 13, 29, 40

includes Computer Problems.



HW # 3

Computer Assignments

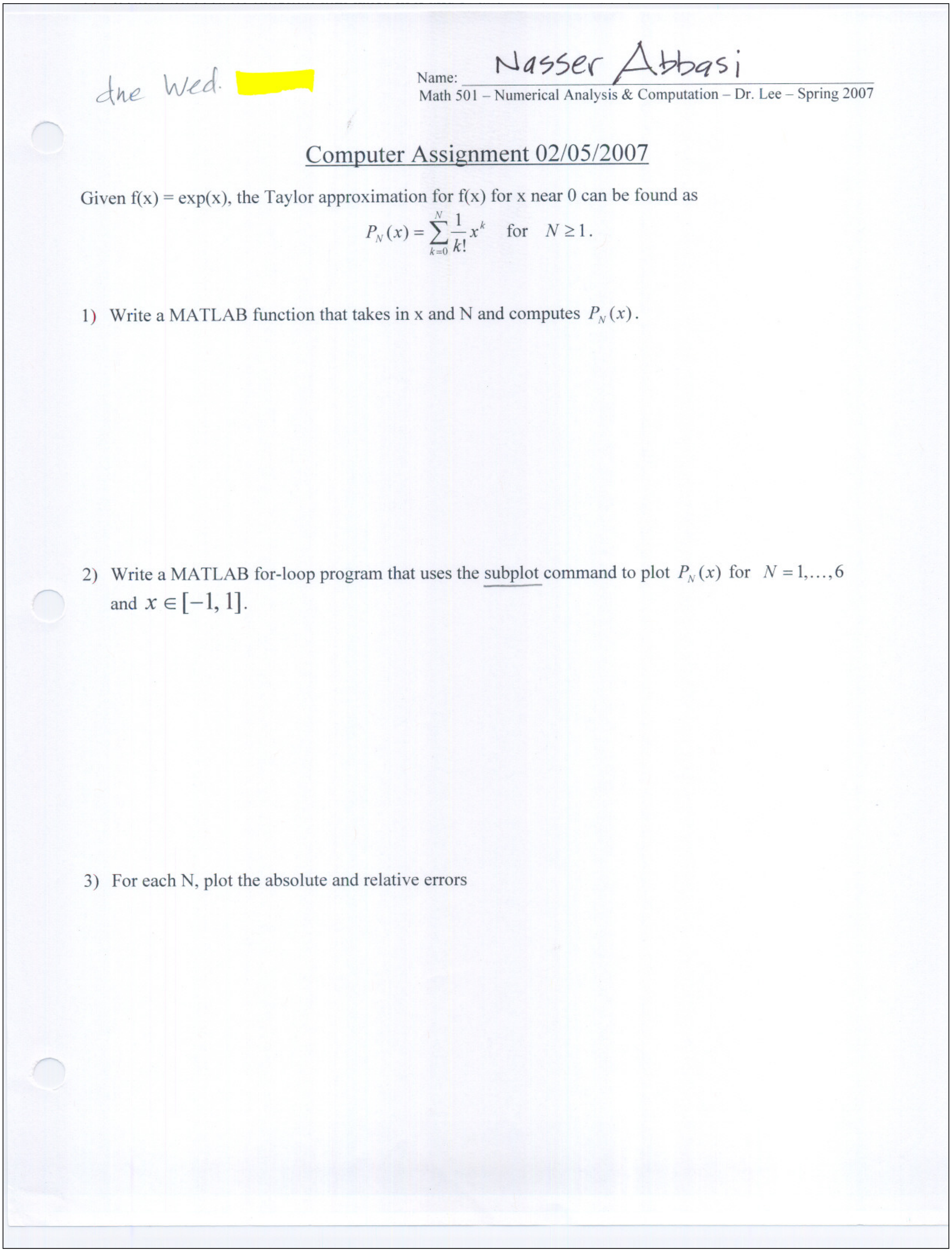
① Taylor approximation 15/15

② Bisection and secant methods for 1-D 10/10

③ Horner Method. 10/10

Note: Matlab Source code if needed can be Found at This temporary folder:

http://12000.org/tmp/021407



due Wed. [redacted]

Name: Nasser Abbasi

Math 501 – Numerical Analysis & Computation – Dr. Lee – Spring 2007

Computer Assignment 02/05/2007

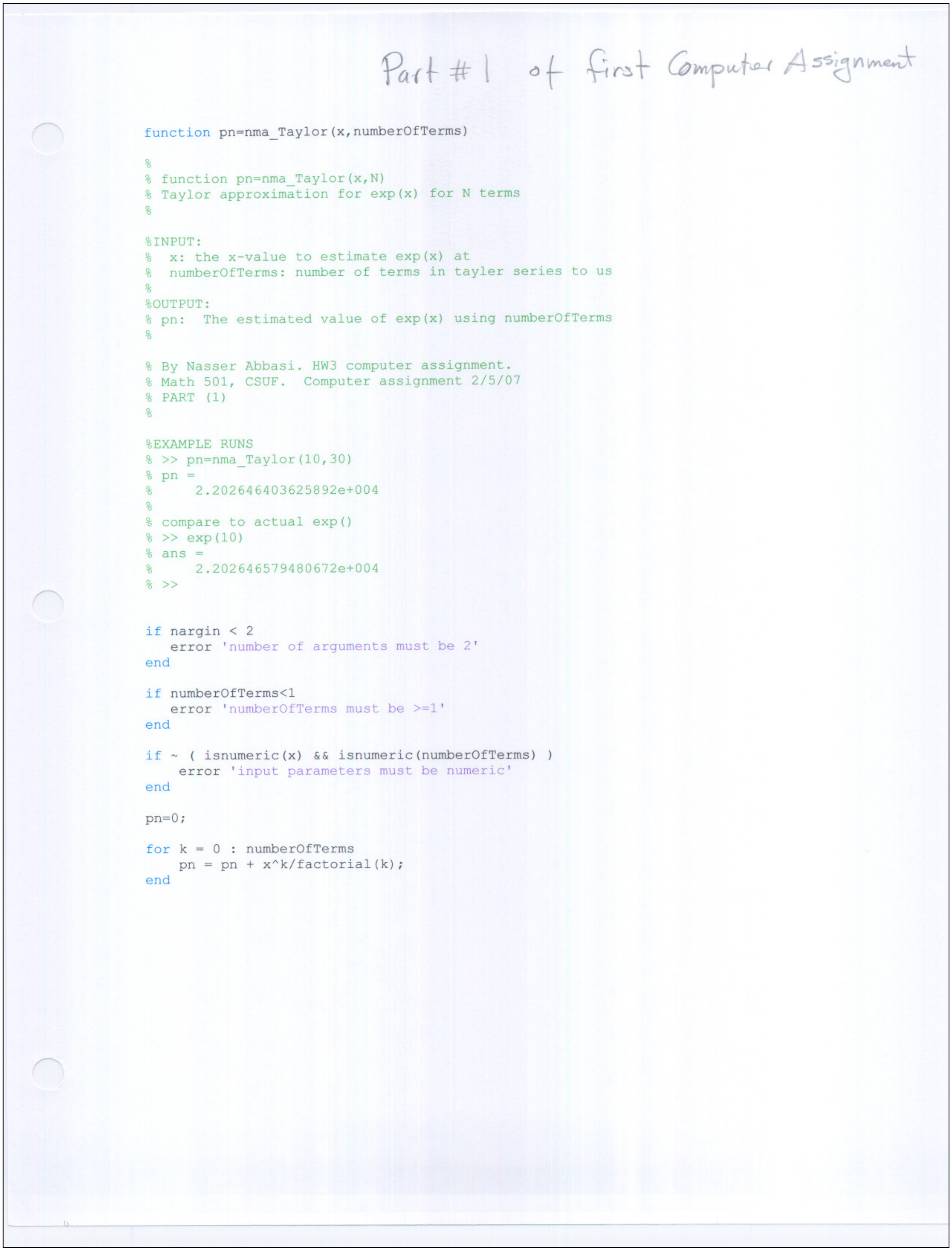
Given f(x) = exp(x), the Taylor approximation for f(x) for x near 0 can be found as

$$P_N(x) = \sum_{k=0}^{N} \frac{1}{k!} x^k \quad \text{for} \quad N \geq 1.$$

1) Write a MATLAB function that takes in x and N and computes $P_N(x)$.

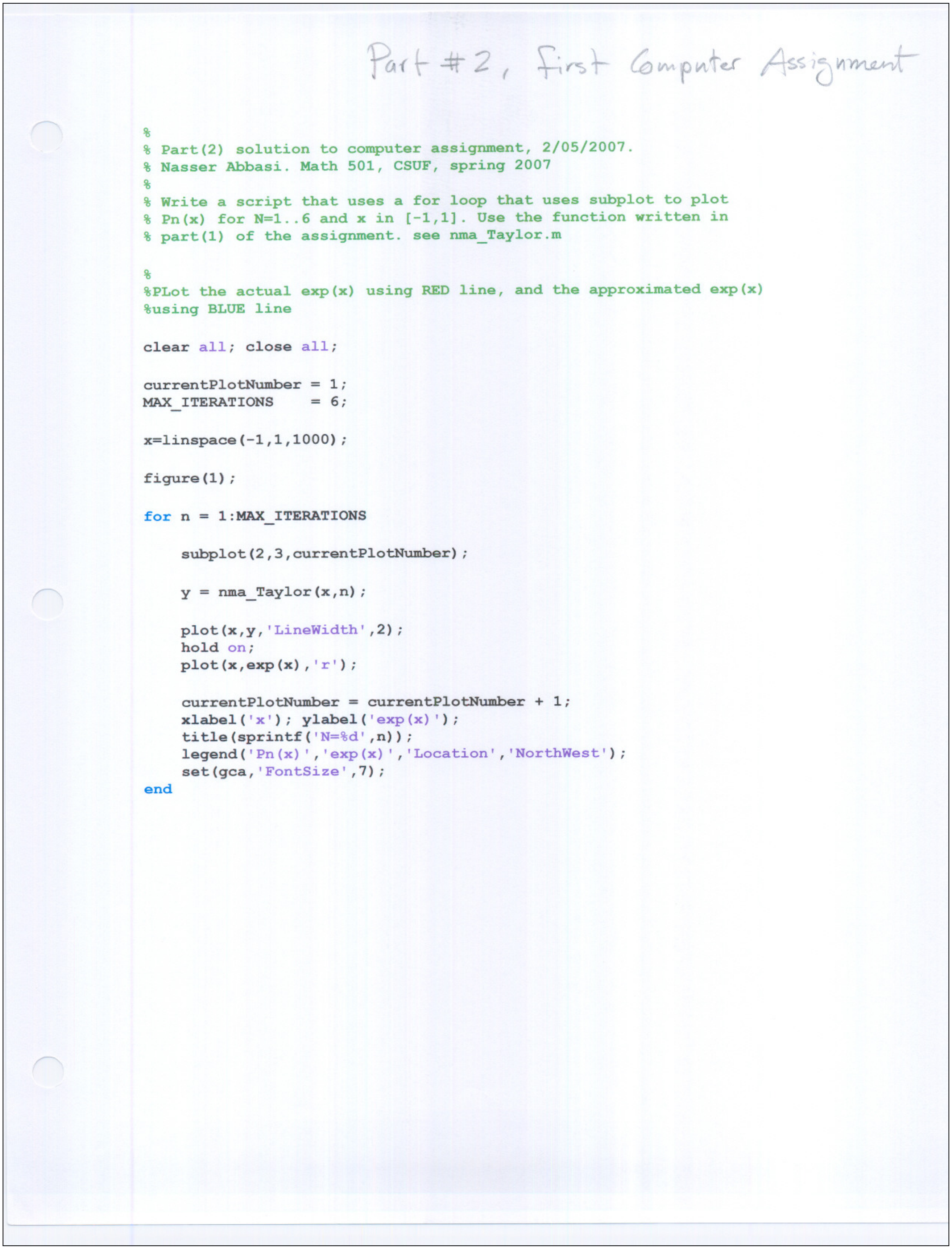
2) Write a MATLAB for-loop program that uses the subplot command to plot $P_N(x)$ for $N = 1, \ldots, 6$ and $x \in [-1, 1]$.

3) For each N, plot the absolute and relative errors



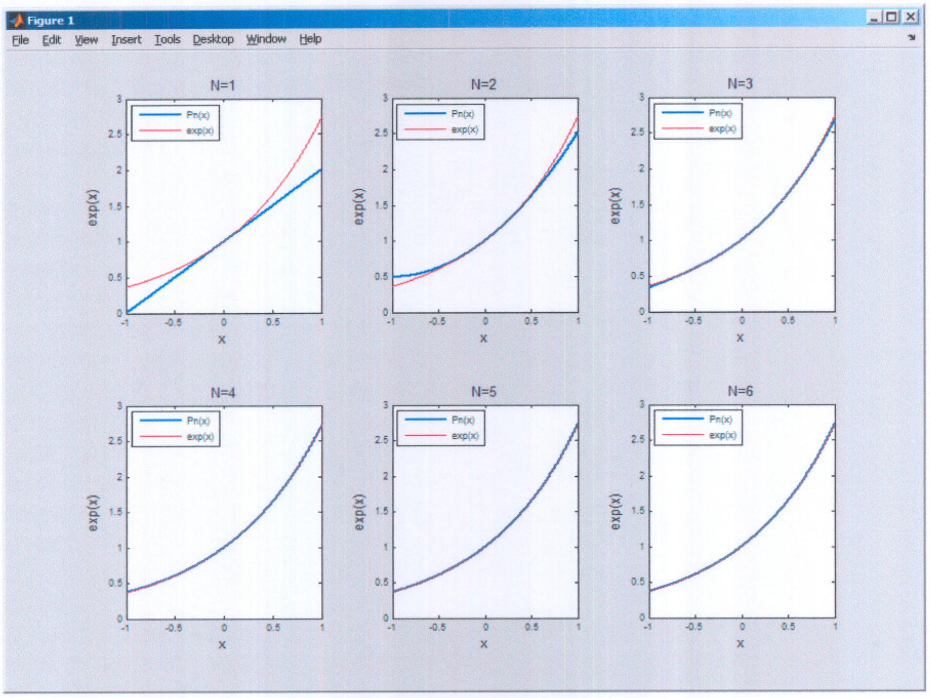
Part # 1 of first Computer Assignment

```
function pn=nma_Taylor(x,numberOfTerms)

%
% function pn=nma_Taylor(x,N)
% Taylor approximation for exp(x) for N terms
%

%INPUT:
%  x: the x-value to estimate exp(x) at
%  numberOfTerms: number of terms in tayler series to us
%
%OUTPUT:
% pn:  The estimated value of exp(x) using numberOfTerms
%

% By Nasser Abbasi. HW3 computer assignment.
% Math 501, CSUF.  Computer assignment 2/5/07
% PART (1)
%

%EXAMPLE RUNS
% >> pn=nma_Taylor(10,30)
% pn =
%      2.202646403625892e+004
%
% compare to actual exp()
% >> exp(10)
% ans =
%      2.202646579480672e+004
% >>


if nargin < 2
   error 'number of arguments must be 2'
end

if numberOfTerms<1
   error 'numberOfTerms must be >=1'
end

if ~ ( isnumeric(x) && isnumeric(numberOfTerms) )
    error 'input parameters must be numeric'
end

pn=0;

for k = 0 : numberOfTerms
    pn = pn + x^k/factorial(k);
end
```

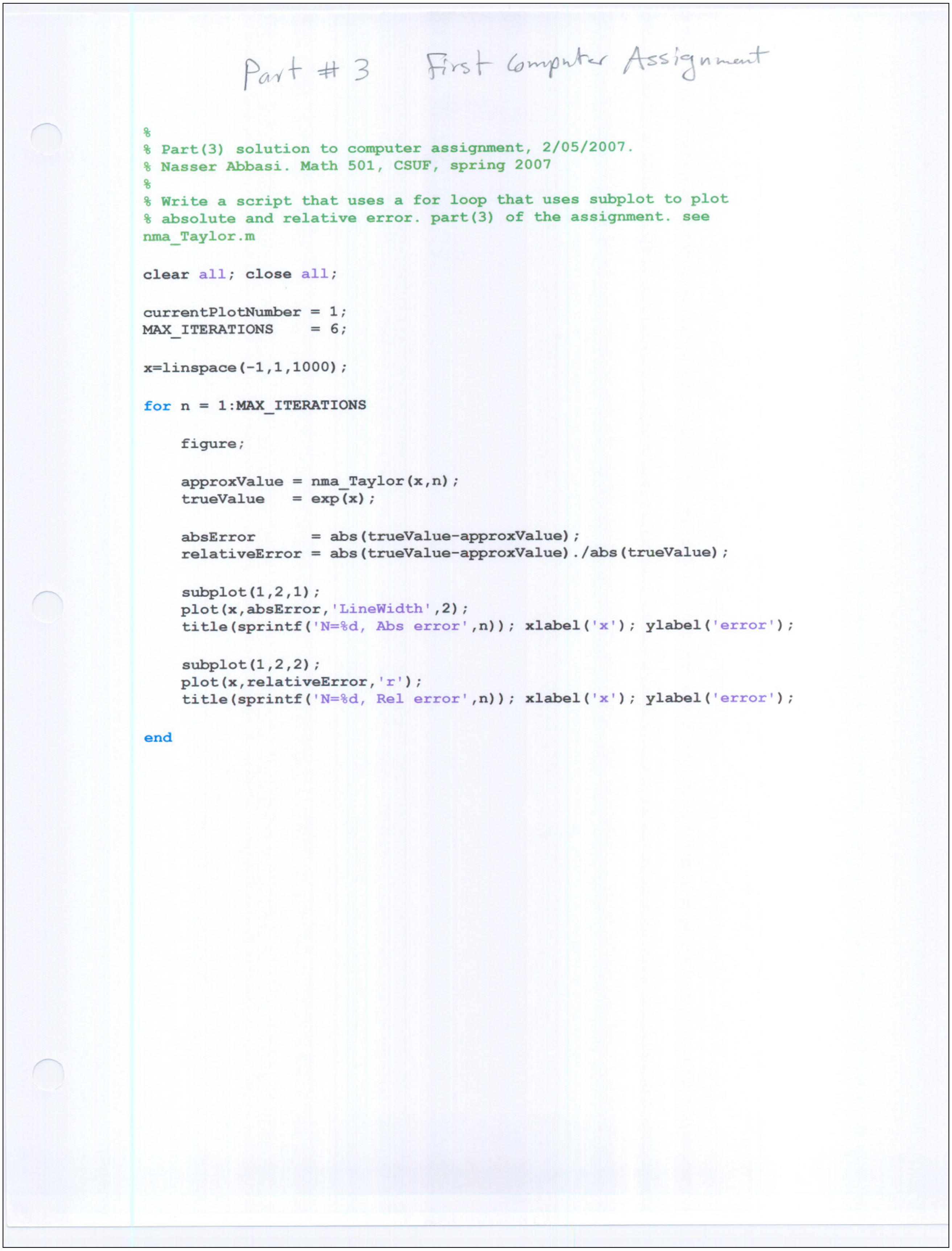


Part #2, first Computer Assignment

```
%
% Part(2) solution to computer assignment, 2/05/2007.
% Nasser Abbasi. Math 501, CSUF, spring 2007
%
% Write a script that uses a for loop that uses subplot to plot
% Pn(x) for N=1..6 and x in [-1,1]. Use the function written in
% part(1) of the assignment. see nma_Taylor.m

%
%PLot the actual exp(x) using RED line, and the approximated exp(x)
%using BLUE line

clear all; close all;

currentPlotNumber = 1;
MAX_ITERATIONS    = 6;

x=linspace(-1,1,1000);

figure(1);

for n = 1:MAX_ITERATIONS

    subplot(2,3,currentPlotNumber);

    y = nma_Taylor(x,n);

    plot(x,y,'LineWidth',2);
    hold on;
    plot(x,exp(x),'r');

    currentPlotNumber = currentPlotNumber + 1;
    xlabel('x'); ylabel('exp(x)');
    title(sprintf('N=%d',n));
    legend('Pn(x)','exp(x)','Location','NorthWest');
    set(gca,'FontSize',7);
end
```



Notice: As N increases, approximation (blue Color) approaches actual value (Red Color).
at N=6 There is almost No visible difference.



Part # 3 First Computer Assignment

```
%
% Part(3) solution to computer assignment, 2/05/2007.
% Nasser Abbasi. Math 501, CSUF, spring 2007
%
% Write a script that uses a for loop that uses subplot to plot
% absolute and relative error. part(3) of the assignment. see
nma_Taylor.m

clear all; close all;

currentPlotNumber = 1;
MAX_ITERATIONS    = 6;

x=linspace(-1,1,1000);

for n = 1:MAX_ITERATIONS

    figure;

    approxValue = nma_Taylor(x,n);
    trueValue   = exp(x);

    absError      = abs(trueValue-approxValue);
    relativeError = abs(trueValue-approxValue)./abs(trueValue);

    subplot(1,2,1);
    plot(x,absError,'LineWidth',2);
    title(sprintf('N=%d, Abs error',n)); xlabel('x'); ylabel('error');

    subplot(1,2,2);
    plot(x,relativeError,'r');
    title(sprintf('N=%d, Rel error',n)); xlabel('x'); ylabel('error');

end
```

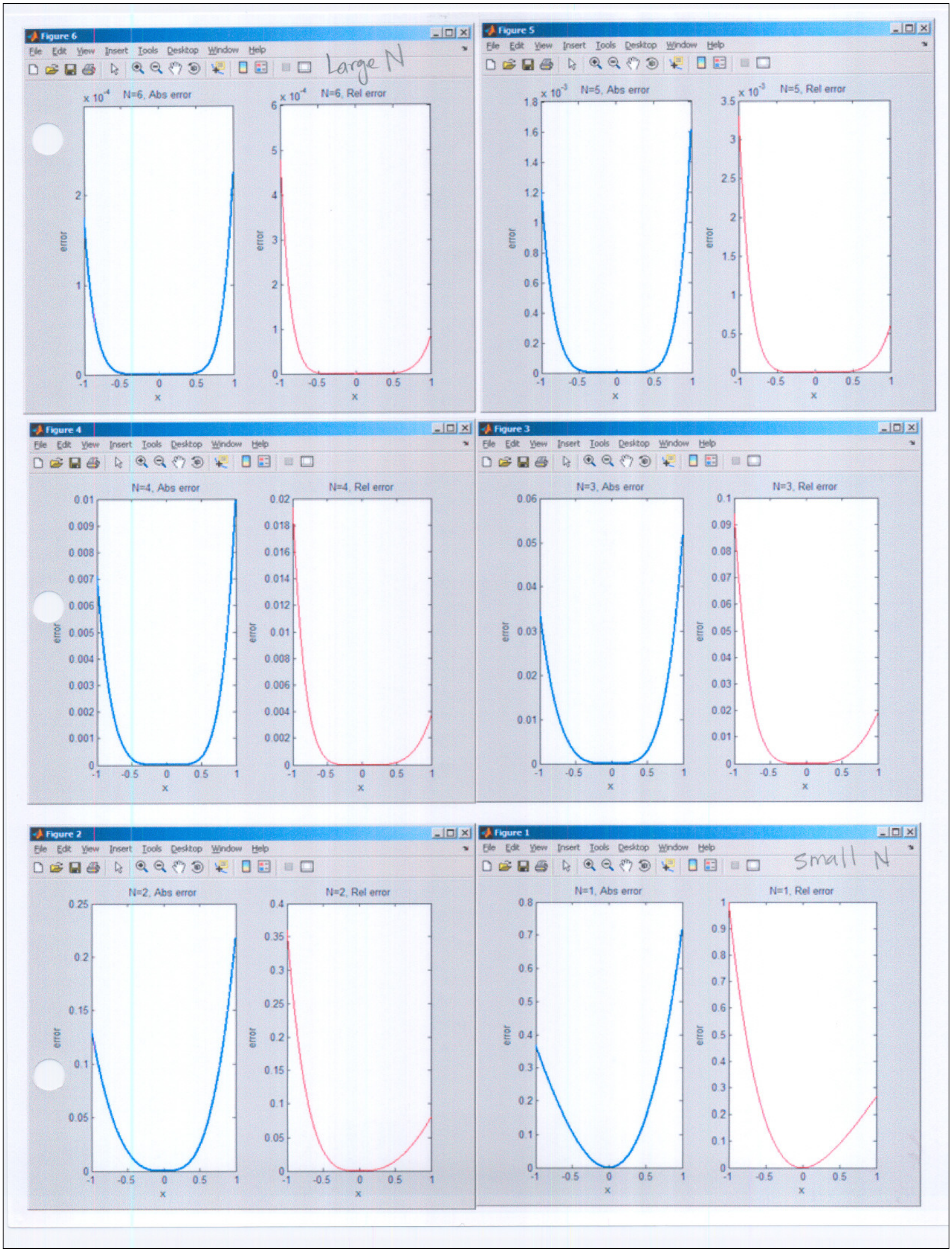
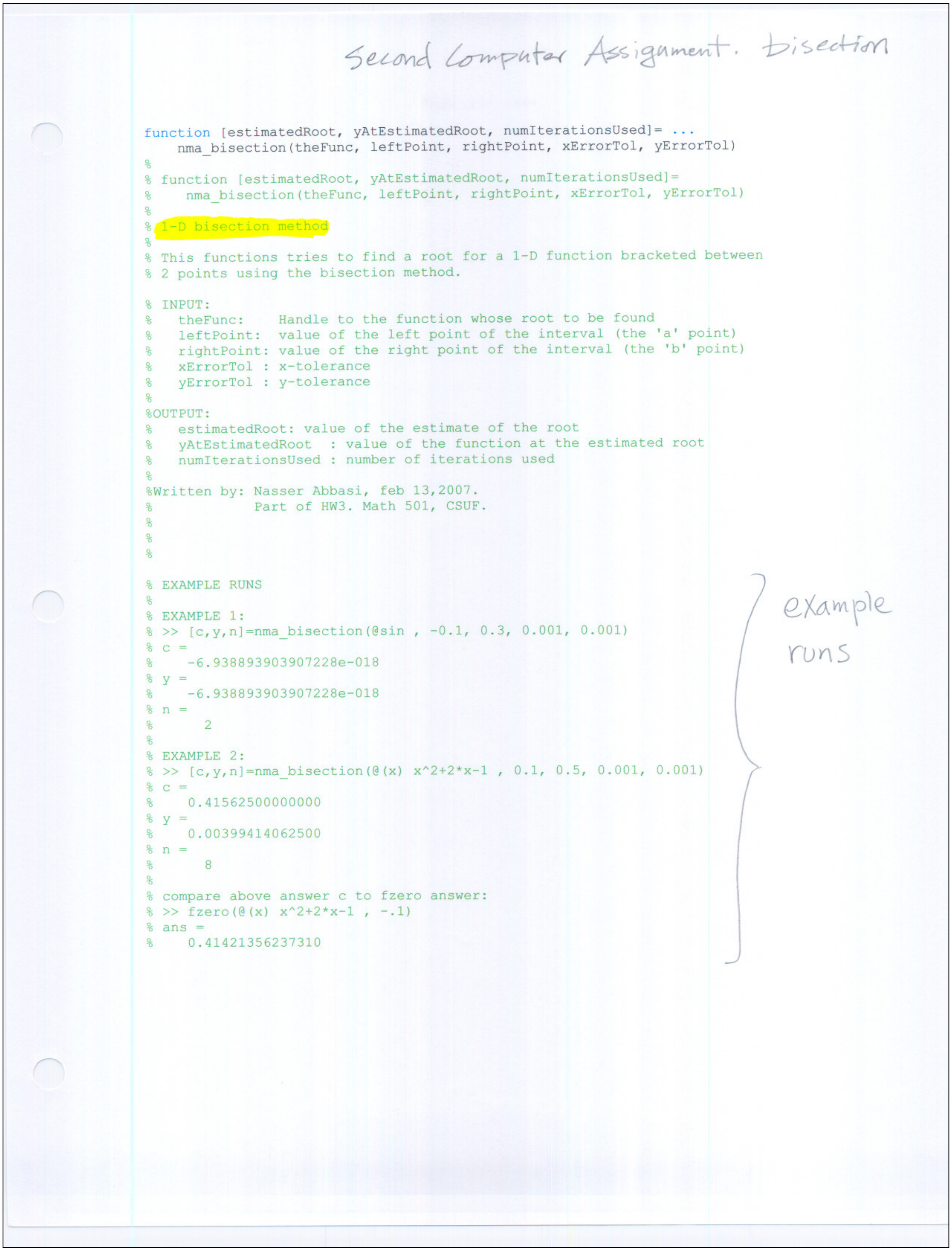


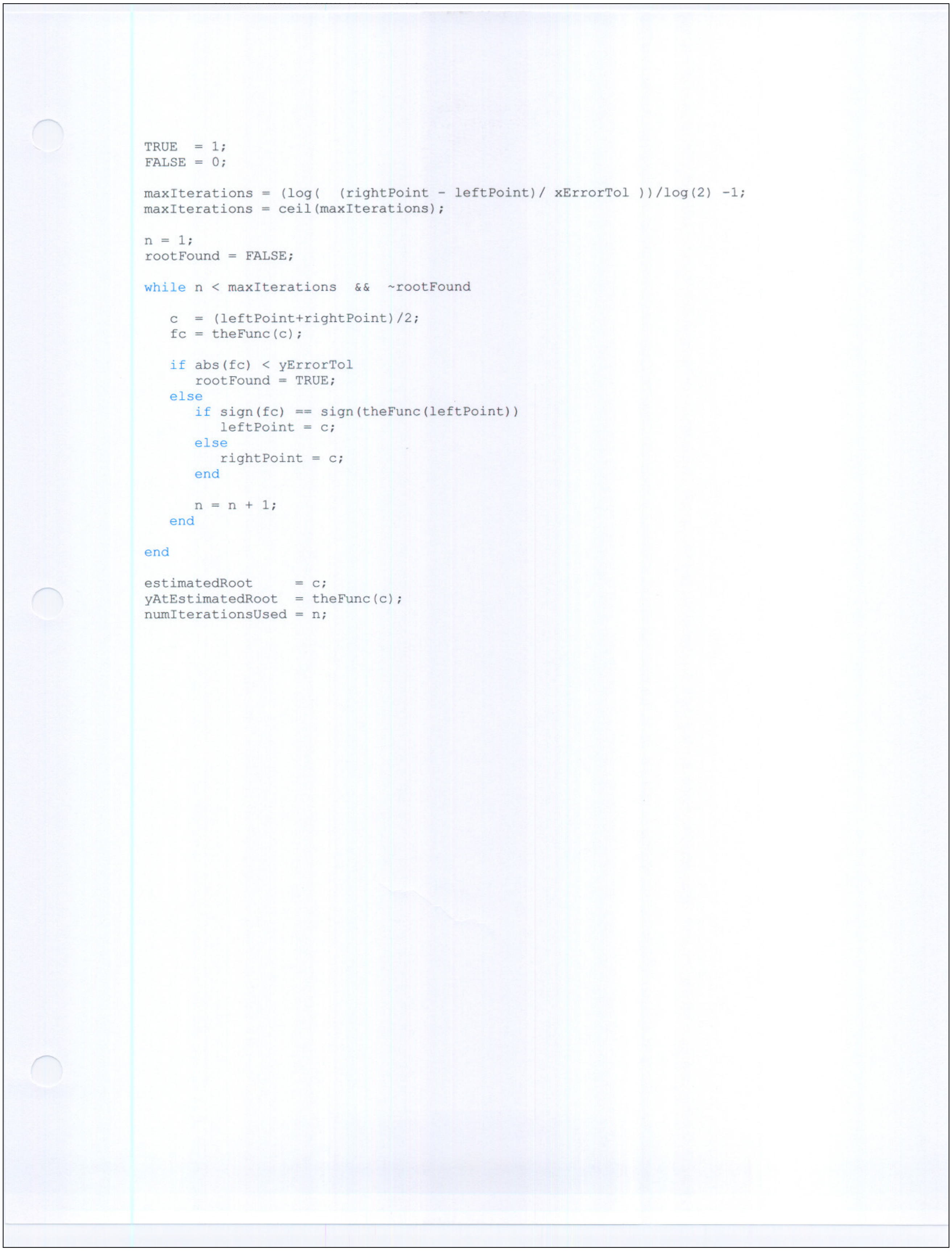
Figure 6
File Edit View Insert Tools Desktop Window Help
Large N
x 10^-4 N=6, Abs error
0
1
2
error
-1 -0.5 0 0.5 1
x
x 10^-4 N=6, Rel error
0
1
2
3
4
5
6
error
-1 -0.5 0 0.5 1
x
Figure 5
File Edit View Insert Tools Desktop Window Help
x 10^-3 N=5, Abs error
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
error
-1 -0.5 0 0.5 1
x
x 10^-3 N=5, Rel error
0
0.5
1
1.5
2
2.5
3
3.5
error
-1 -0.5 0 0.5 1
x
Figure 4
File Edit View Insert Tools Desktop Window Help
N=4, Abs error
0
0.001
0.002
0.003
0.004
0.005
0.006
0.007
0.008
0.009
0.01
error
-1 -0.5 0 0.5 1
x
N=4, Rel error
0
0.002
0.004
0.006
0.008
0.01
0.012
0.014
0.016
0.018
0.02
error
-1 -0.5 0 0.5 1
x
Figure 3
File Edit View Insert Tools Desktop Window Help
N=3, Abs error
0
0.01
0.02
0.03
0.04
0.05
0.06
error
-1 -0.5 0 0.5 1
x
N=3, Rel error
0
0.01
0.02
0.03
0.04
0.05
0.06
0.07
0.08
0.09
0.1
error
-1 -0.5 0 0.5 1
x
Figure 2
File Edit View Insert Tools Desktop Window Help
N=2, Abs error
0
0.05
0.1
0.15
0.2
0.25
error
-1 -0.5 0 0.5 1
x
N=2, Rel error
0
0.05
0.1
0.15
0.2
0.25
0.3
0.35
0.4
error
-1 -0.5 0 0.5 1
x
Figure 1
File Edit View Insert Tools Desktop Window Help
Small N
N=1, Abs error
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
error
-1 -0.5 0 0.5 1
x
N=1, Rel error
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
error
-1 -0.5 0 0.5 1
x



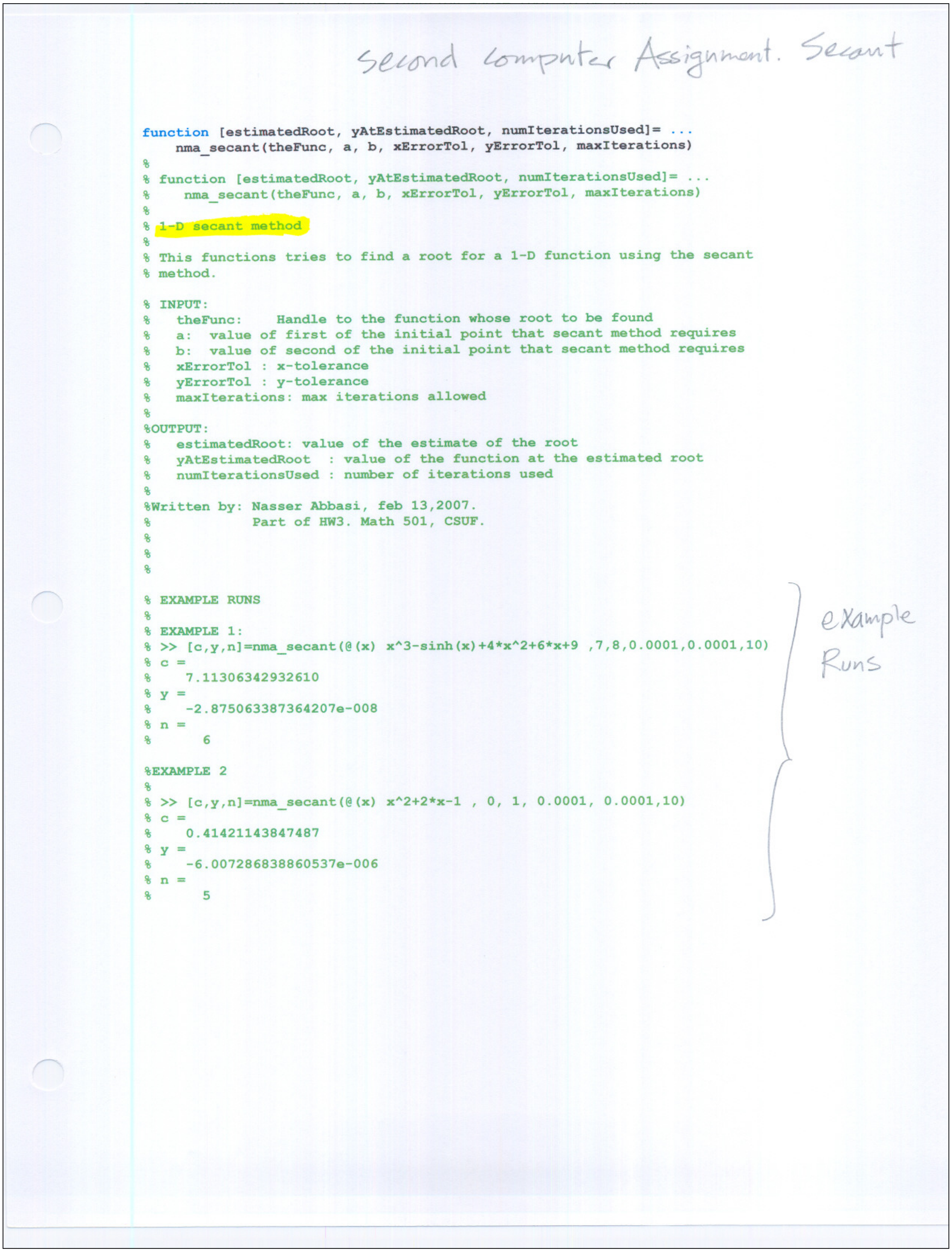
Second Computer Assignment. bisection

```
function [estimatedRoot, yAtEstimatedRoot, numIterationsUsed]= ...
    nma_bisection(theFunc, leftPoint, rightPoint, xErrorTol, yErrorTol)
%
% function [estimatedRoot, yAtEstimatedRoot, numIterationsUsed]=
%    nma_bisection(theFunc, leftPoint, rightPoint, xErrorTol, yErrorTol)
%
% 1-D bisection method
%
% This functions tries to find a root for a 1-D function bracketed between
% 2 points using the bisection method.

% INPUT:
%   theFunc:     Handle to the function whose root to be found
%   leftPoint:  value of the left point of the interval (the 'a' point)
%   rightPoint: value of the right point of the interval (the 'b' point)
%   xErrorTol : x-tolerance
%   yErrorTol : y-tolerance
%
%OUTPUT:
%   estimatedRoot: value of the estimate of the root
%   yAtEstimatedRoot  : value of the function at the estimated root
%   numIterationsUsed : number of iterations used
%
%Written by: Nasser Abbasi, feb 13,2007.
%             Part of HW3. Math 501, CSUF.
%
%
%


% EXAMPLE RUNS
%
% EXAMPLE 1:
% >> [c,y,n]=nma_bisection(@sin , -0.1, 0.3, 0.001, 0.001)
% c =
%    -6.938893903907228e-018
% y =
%    -6.938893903907228e-018
% n =
%      2
%
% EXAMPLE 2:
% >> [c,y,n]=nma_bisection(@(x) x^2+2*x-1 , 0.1, 0.5, 0.001, 0.001)
% c =
%    0.41562500000000
% y =
%    0.00399414062500
% n =
%      8
%
% compare above answer c to fzero answer:
% >> fzero(@(x) x^2+2*x-1 , -.1)
% ans =
%    0.41421356237310
```

example runs



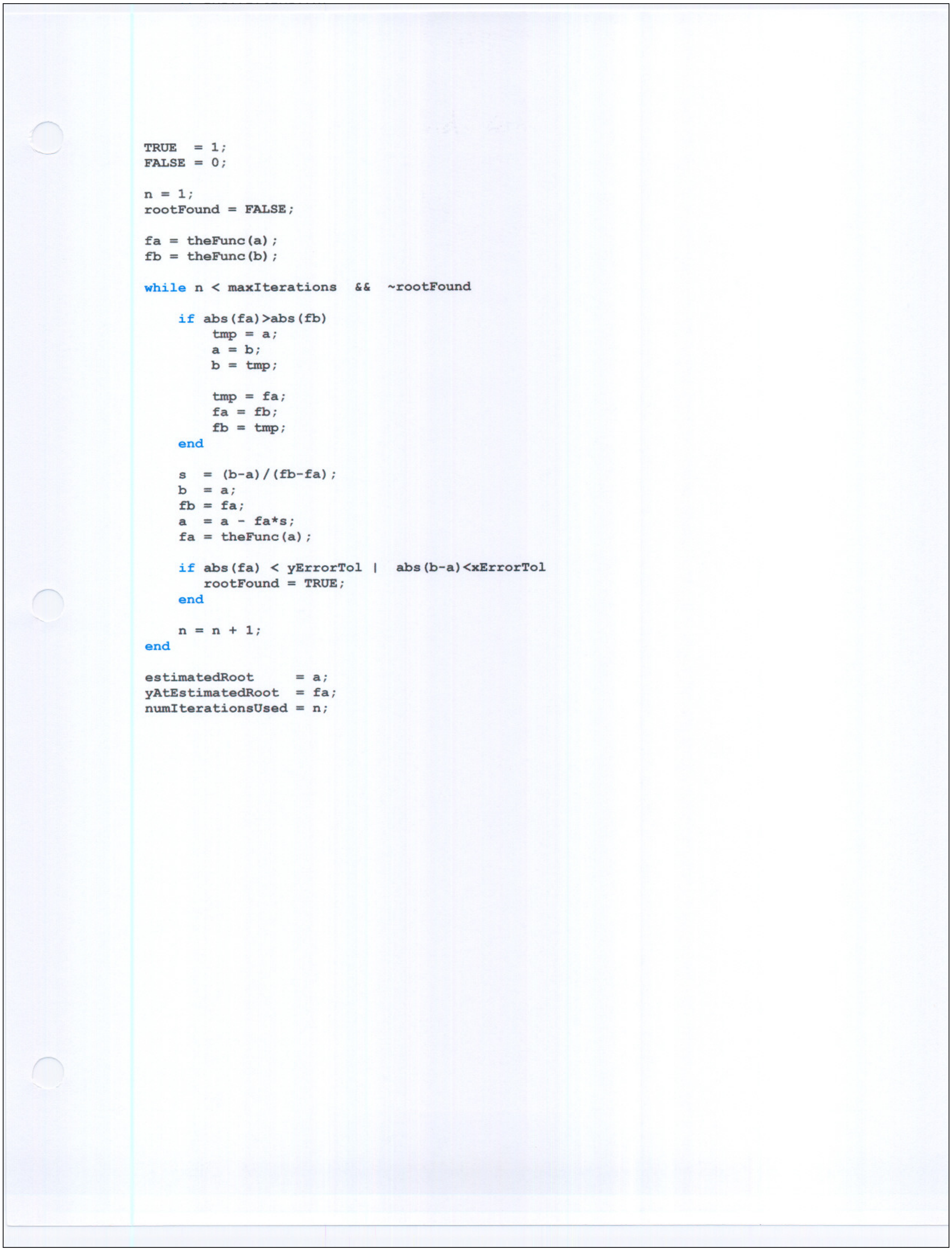
```
TRUE  = 1;
FALSE = 0;

maxIterations = (log(  (rightPoint - leftPoint)/ xErrorTol ))/log(2) -1;
maxIterations = ceil(maxIterations);

n = 1;
rootFound = FALSE;

while n < maxIterations  &&  ~rootFound

   c  = (leftPoint+rightPoint)/2;
   fc = theFunc(c);

   if abs(fc) < yErrorTol
      rootFound = TRUE;
   else
      if sign(fc) == sign(theFunc(leftPoint))
         leftPoint = c;
      else
         rightPoint = c;
      end

      n = n + 1;
   end

end

estimatedRoot     = c;
yAtEstimatedRoot  = theFunc(c);
numIterationsUsed = n;
```



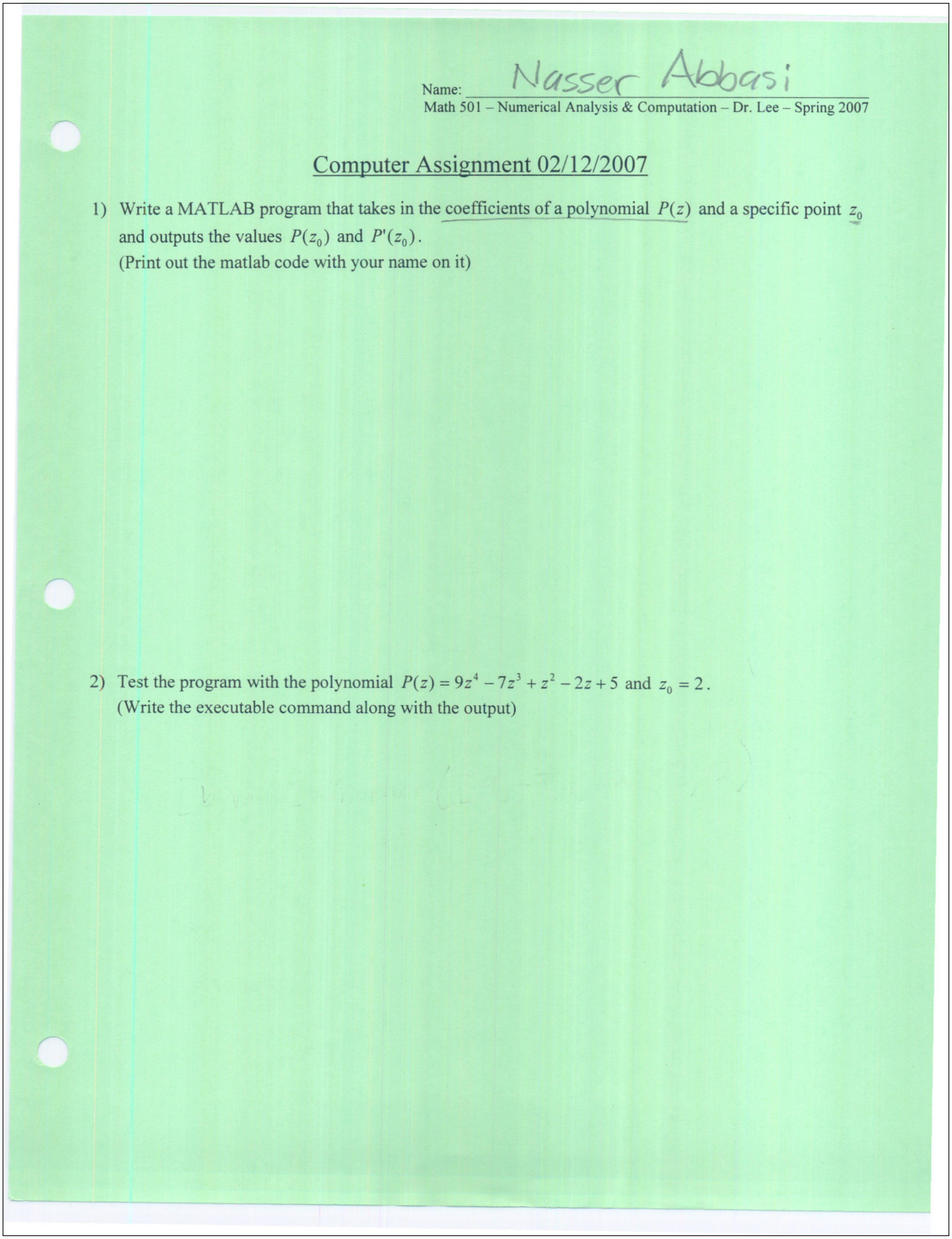
second computer Assignment. Secant

```
function [estimatedRoot, yAtEstimatedRoot, numIterationsUsed]= ...
    nma_secant(theFunc, a, b, xErrorTol, yErrorTol, maxIterations)
%
% function [estimatedRoot, yAtEstimatedRoot, numIterationsUsed]= ...
%    nma_secant(theFunc, a, b, xErrorTol, yErrorTol, maxIterations)
%
% 1-D secant method
%
% This functions tries to find a root for a 1-D function using the secant
% method.

% INPUT:
%   theFunc:    Handle to the function whose root to be found
%   a:  value of first of the initial point that secant method requires
%   b:  value of second of the initial point that secant method requires
%   xErrorTol : x-tolerance
%   yErrorTol : y-tolerance
%   maxIterations: max iterations allowed
%
%OUTPUT:
%   estimatedRoot: value of the estimate of the root
%   yAtEstimatedRoot  : value of the function at the estimated root
%   numIterationsUsed : number of iterations used
%
%Written by: Nasser Abbasi, feb 13,2007.
%            Part of HW3. Math 501, CSUF.
%
%
%

% EXAMPLE RUNS
%
% EXAMPLE 1:
% >> [c,y,n]=nma_secant(@(x) x^3-sinh(x)+4*x^2+6*x+9 ,7,8,0.0001,0.0001,10)
% c =
%    7.11306342932610
% y =
%    -2.875063387364207e-008
% n =
%      6

%EXAMPLE 2
%
% >> [c,y,n]=nma_secant(@(x) x^2+2*x-1 , 0, 1, 0.0001, 0.0001,10)
% c =
%    0.41421143847487
% y =
%    -6.007286838860537e-006
% n =
%      5
```

example Runs



```
TRUE  = 1;
FALSE = 0;

n = 1;
rootFound = FALSE;

fa = theFunc(a);
fb = theFunc(b);

while n < maxIterations  &&  ~rootFound

    if abs(fa)>abs(fb)
        tmp = a;
        a = b;
        b = tmp;

        tmp = fa;
        fa = fb;
        fb = tmp;
    end

    s  = (b-a)/(fb-fa);
    b  = a;
    fb = fa;
    a  = a - fa*s;
    fa = theFunc(a);

    if abs(fa) < yErrorTol |  abs(b-a)<xErrorTol
       rootFound = TRUE;
    end

    n = n + 1;
end

estimatedRoot     = a;
yAtEstimatedRoot  = fa;
numIterationsUsed = n;
```



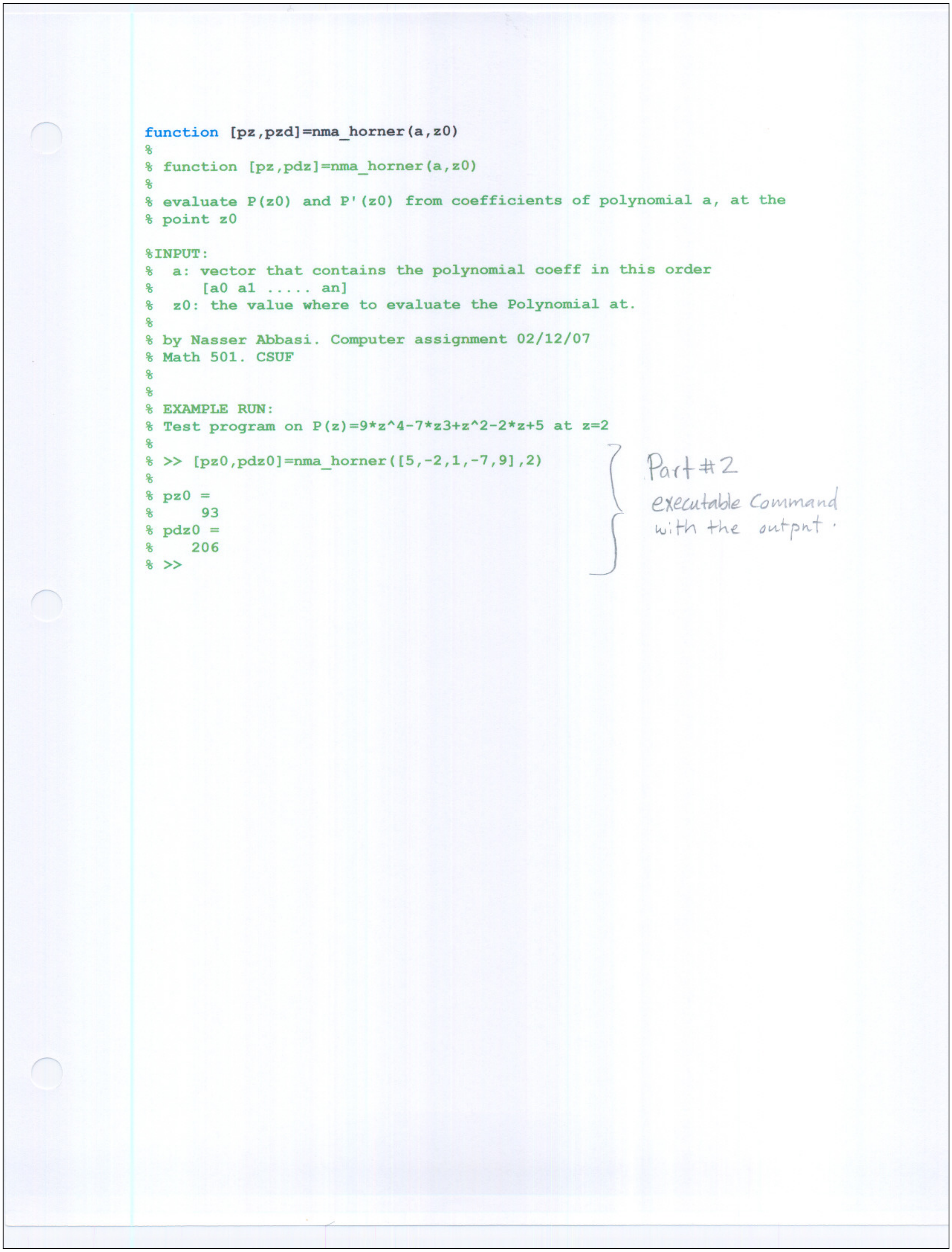
Name: Nasser Abbasi

Math 501 – Numerical Analysis & Computation – Dr. Lee – Spring 2007

Computer Assignment 02/12/2007

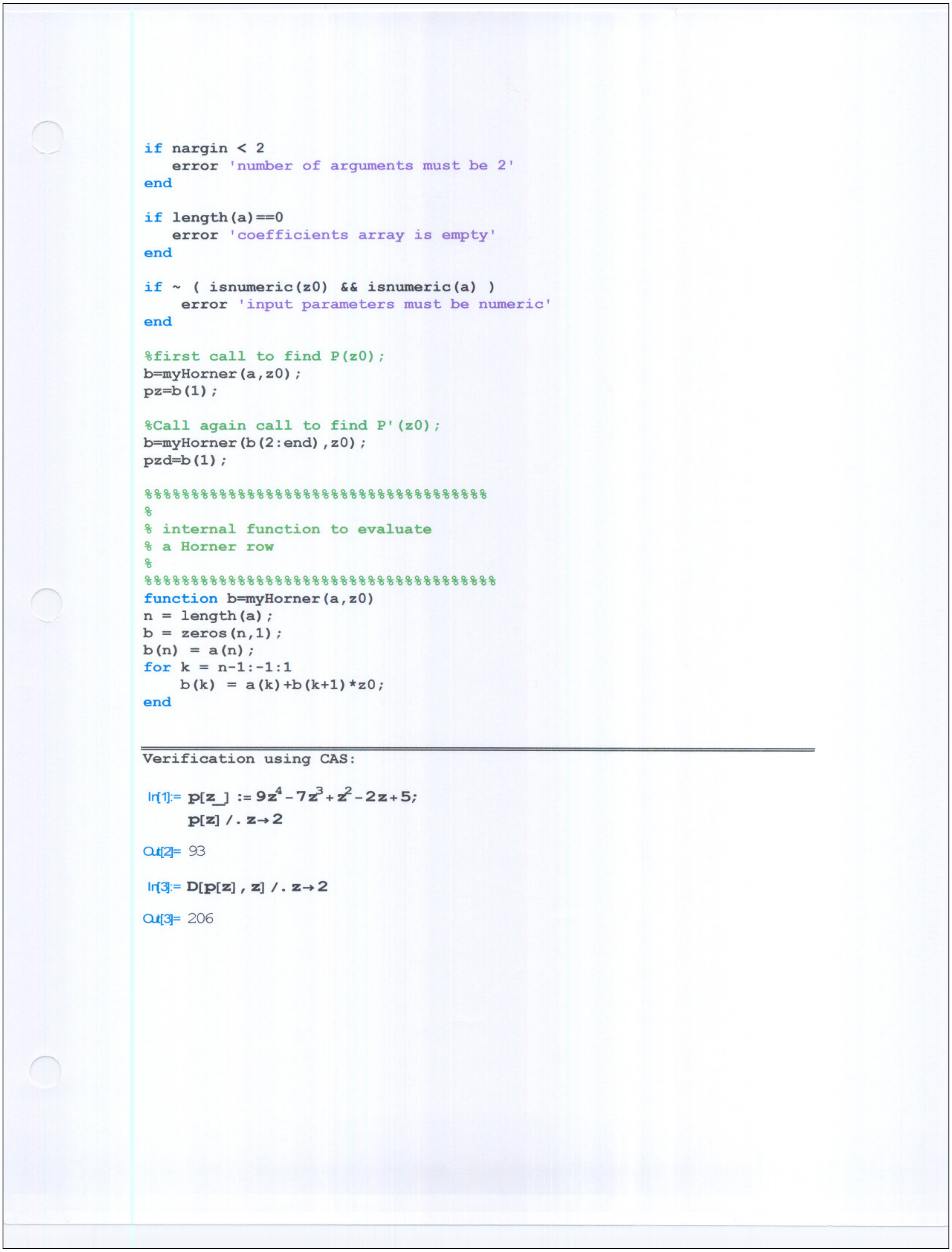
1) Write a MATLAB program that takes in the coefficients of a polynomial $P(z)$ and a specific point z_0 and outputs the values $P(z_0)$ and $P'(z_0)$.
(Print out the matlab code with your name on it)

2) Test the program with the polynomial $P(z) = 9z^4 - 7z^3 + z^2 - 2z + 5$ and $z_0 = 2$.
(Write the executable command along with the output)



```
function [pz,pzd]=nma_horner(a,z0)
%
% function [pz,pdz]=nma_horner(a,z0)
%
% evaluate P(z0) and P'(z0) from coefficients of polynomial a, at the
% point z0

%INPUT:
%  a: vector that contains the polynomial coeff in this order
%     [a0 a1 ..... an]
%  z0: the value where to evaluate the Polynomial at.
%
% by Nasser Abbasi. Computer assignment 02/12/07
% Math 501. CSUF
%
%
% EXAMPLE RUN:
% Test program on P(z)=9*z^4-7*z3+z^2-2*z+5 at z=2
%
% >> [pz0,pdz0]=nma_horner([5,-2,1,-7,9],2)
%
% pz0 =
%     93
% pdz0 =
%    206
% >>
```

Part #2
executable Command
with the output.

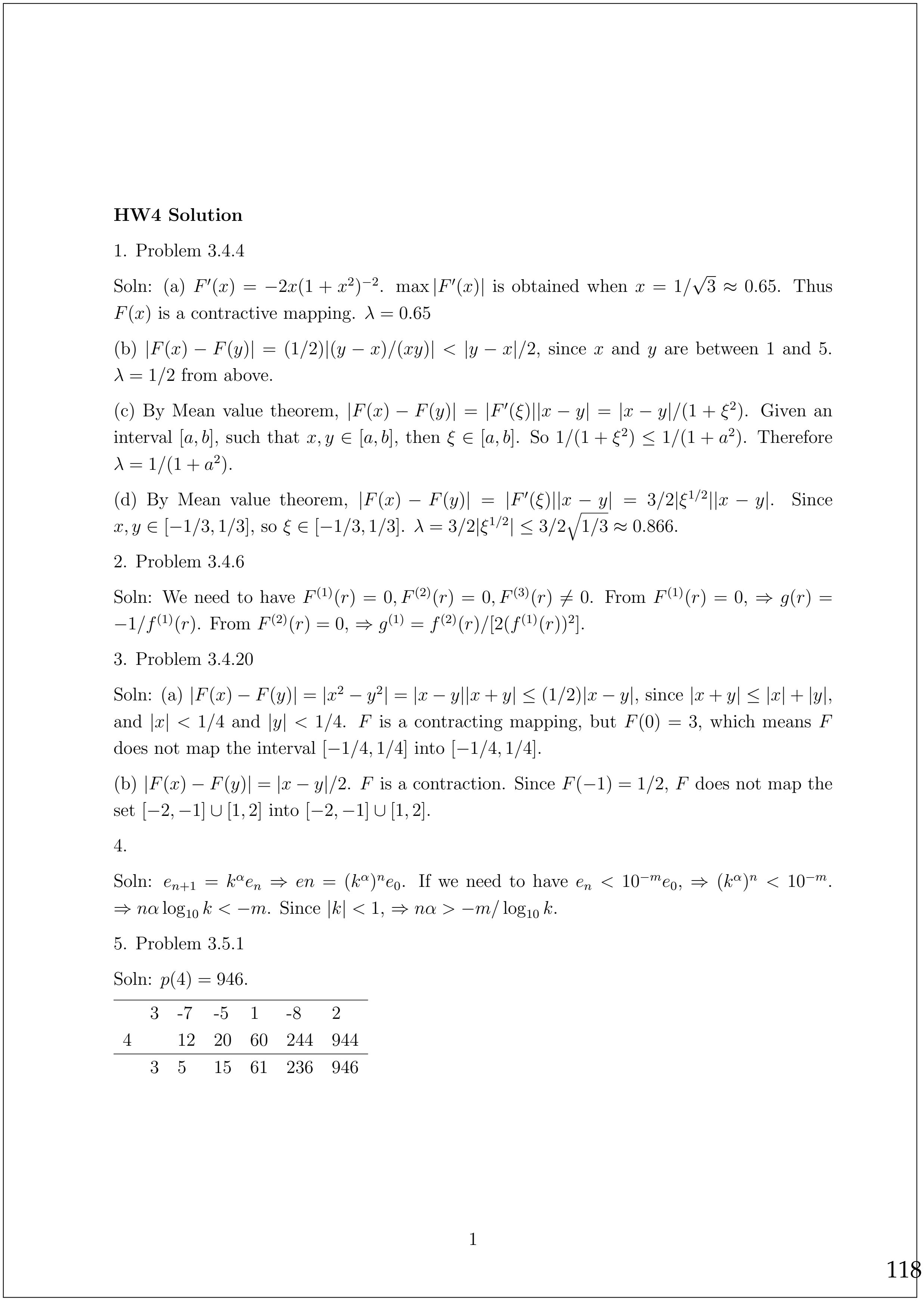


```
if nargin < 2
   error 'number of arguments must be 2'
end

if length(a)==0
   error 'coefficients array is empty'
end

if ~ ( isnumeric(z0) && isnumeric(a) )
    error 'input parameters must be numeric'
end

%first call to find P(z0);
b=myHorner(a,z0);
pz=b(1);

%Call again call to find P'(z0);
b=myHorner(b(2:end),z0);
pzd=b(1);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%
% internal function to evaluate
% a Horner row
%
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function b=myHorner(a,z0)
n = length(a);
b = zeros(n,1);
b(n) = a(n);
for k = n-1:-1:1
    b(k) = a(k)+b(k+1)*z0;
end
```

Verification using CAS:

```
In[1]:= p[z_] := 9z^4-7z^3+z^2-2z+5;
        p[z] /. z→2

Out[2]= 93

In[3]:= D[p[z], z] /. z→2

Out[3]= 206
```

4.4.3 Key solution



HW4 Solution

1. Problem 3.4.4

Soln: (a) $F'(x) = -2x(1+x^2)^{-2}$. $\max|F'(x)|$ is obtained when $x = 1/\sqrt{3} \approx 0.65$. Thus $F(x)$ is a contractive mapping. $\lambda = 0.65$

(b) $|F(x) - F(y)| = (1/2)|(y-x)/(xy)| < |y-x|/2$, since x and y are between 1 and 5. $\lambda = 1/2$ from above.

(c) By Mean value theorem, $|F(x) - F(y)| = |F'(\xi)||x-y| = |x-y|/(1+\xi^2)$. Given an interval $[a,b]$, such that $x, y \in [a,b]$, then $\xi \in [a,b]$. So $1/(1+\xi^2) \leq 1/(1+a^2)$. Therefore $\lambda = 1/(1+a^2)$.

(d) By Mean value theorem, $|F(x) - F(y)| = |F'(\xi)||x-y| = 3/2|\xi^{1/2}||x-y|$. Since $x, y \in [-1/3, 1/3]$, so $\xi \in [-1/3, 1/3]$. $\lambda = 3/2|\xi^{1/2}| \leq 3/2\sqrt{1/3} \approx 0.866$.

2. Problem 3.4.6

Soln: We need to have $F^{(1)}(r) = 0, F^{(2)}(r) = 0, F^{(3)}(r) \neq 0$. From $F^{(1)}(r) = 0, \Rightarrow g(r) = -1/f^{(1)}(r)$. From $F^{(2)}(r) = 0, \Rightarrow g^{(1)} = f^{(2)}(r)/[2(f^{(1)}(r))^2]$.

3. Problem 3.4.20

Soln: (a) $|F(x) - F(y)| = |x^2 - y^2| = |x-y||x+y| \leq (1/2)|x-y|$, since $|x+y| \leq |x| + |y|$, and $|x| < 1/4$ and $|y| < 1/4$. F is a contracting mapping, but $F(0) = 3$, which means F does not map the interval $[-1/4, 1/4]$ into $[-1/4, 1/4]$.

(b) $|F(x) - F(y)| = |x-y|/2$. F is a contraction. Since $F(-1) = 1/2$, F does not map the set $[-2,-1] \cup [1,2]$ into $[-2,-1] \cup [1,2]$.

4.

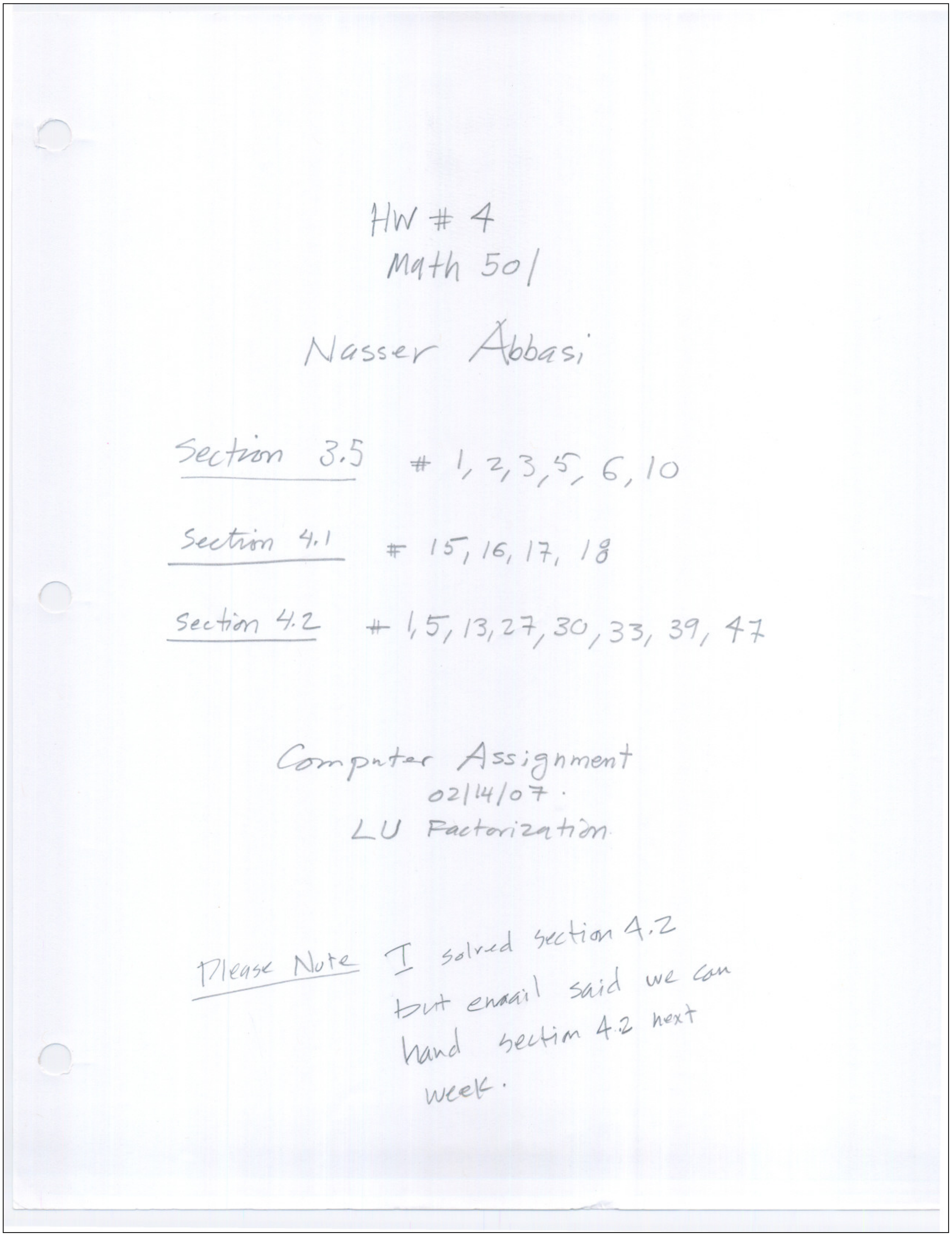
Soln: $e_{n+1} = k^\alpha e_n \Rightarrow en = (k^\alpha)^n e_0$. If we need to have $e_n < 10^{-m} e_0$, $\Rightarrow (k^\alpha)^n < 10^{-m}$. $\Rightarrow n\alpha \log_{10} k < -m$. Since $|k| < 1$, $\Rightarrow n\alpha > -m/\log_{10} k$.

5. Problem 3.5.1

Soln: $p(4) = 946$.

	3	-7	-5	1	-8	2
4		12	20	60	244	944
	3	5	15	61	236	946

4.5 HW 4



HW # 4
Math 501

Nasser Abbasi

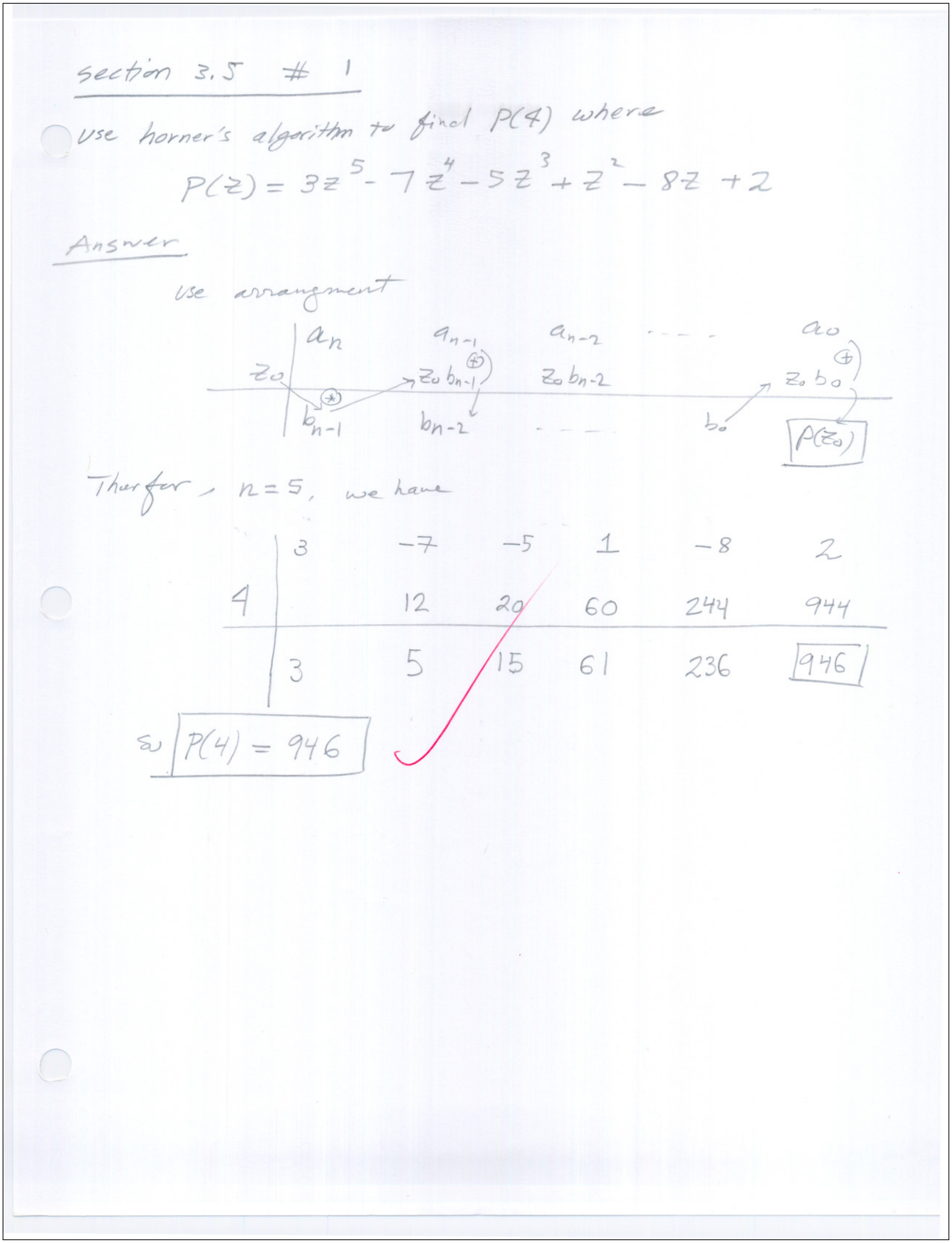
Section 3.5 # 1, 2, 3, 5, 6, 10

Section 4.1 # 15, 16, 17, 18

Section 4.2 # 1, 5, 13, 27, 30, 33, 39, 47

Computer Assignment
02/14/07.
LU Factorization.

Please Note I solved section 4.2 but email said we can hand section 4.2 next week.



section 3.5 # 1

use horner's algorithm to find $P(4)$ where

$$P(z) = 3z^5 - 7z^4 - 5z^3 + z^2 - 8z + 2$$

Answer

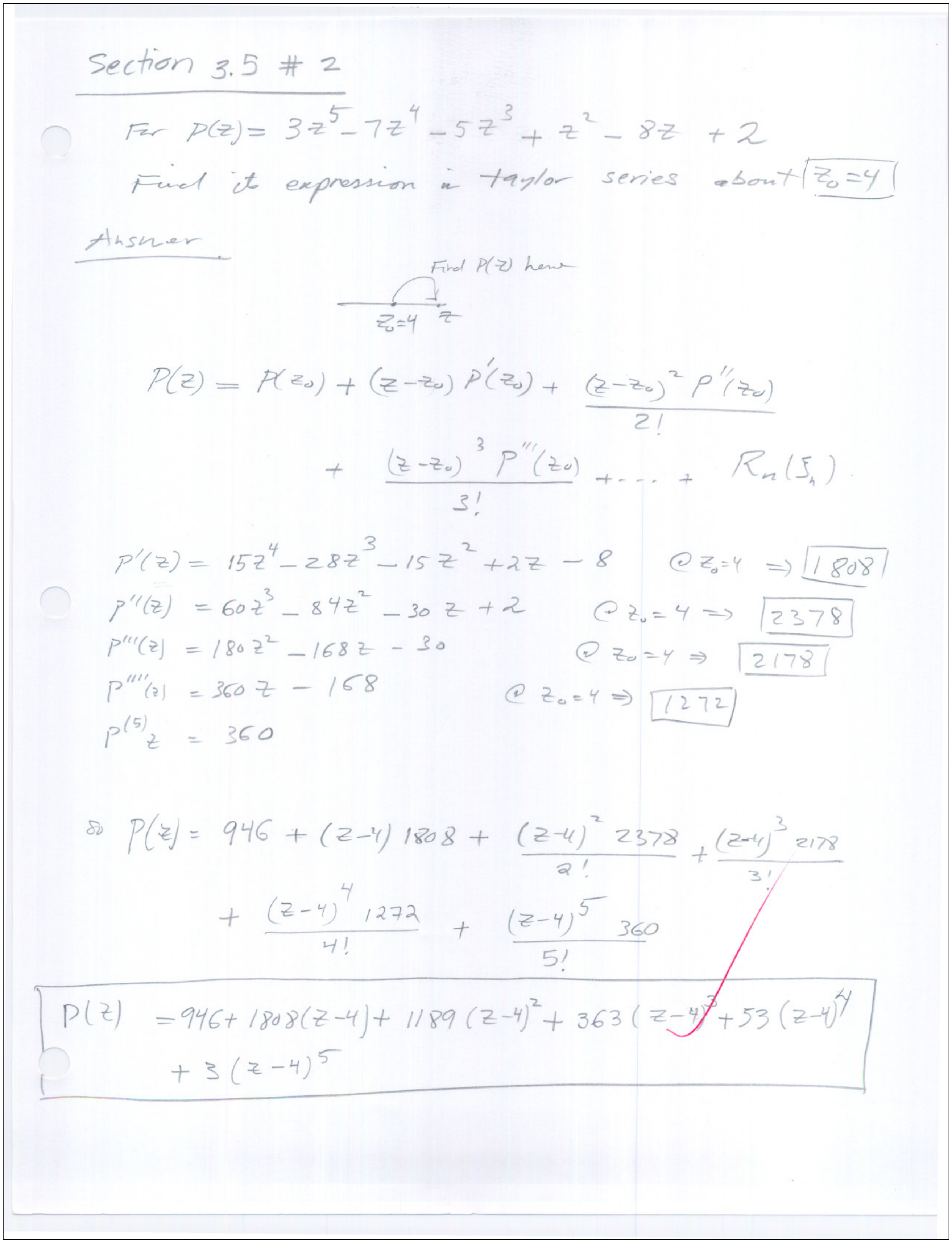
use arrangment

$$\begin{array}{c|ccccc} & a_n & a_{n-1} & a_{n-2} & \cdots & a_0 \\ z_0 & \otimes & z_0 b_{n-1} \oplus & z_0 b_{n-2} & & z_0 b_0 \oplus \\ \hline & b_{n-1} & b_{n-2} & \cdots & b_0 & \boxed{P(z_0)} \end{array}$$

Therfor, $n=5$, we have

$$\begin{array}{c|cccccc} & 3 & -7 & -5 & 1 & -8 & 2 \\ 4 & & 12 & 20 & 60 & 244 & 944 \\ \hline & 3 & 5 & 15 & 61 & 236 & \boxed{946} \end{array}$$

so $\boxed{P(4) = 946}$



Section 3.5 # 2

For $P(z) = 3z^5 - 7z^4 - 5z^3 + z^2 - 8z + 2$

Find it expression in taylor series about $\boxed{z_0=4}$

Answer

Find P(z) here

$z_0=4$ z

$$P(z) = P(z_0) + (z-z_0)P'(z_0) + \frac{(z-z_0)^2 P''(z_0)}{2!} + \frac{(z-z_0)^3 P'''(z_0)}{3!} + \cdots + R_n(\xi_n)$$

$$P'(z) = 15z^4 - 28z^3 - 15z^2 + 2z - 8 \quad @ z_0=4 \Rightarrow \boxed{1808}$$

$$P''(z) = 60z^3 - 84z^2 - 30z + 2 \quad @ z_0=4 \Rightarrow \boxed{2378}$$

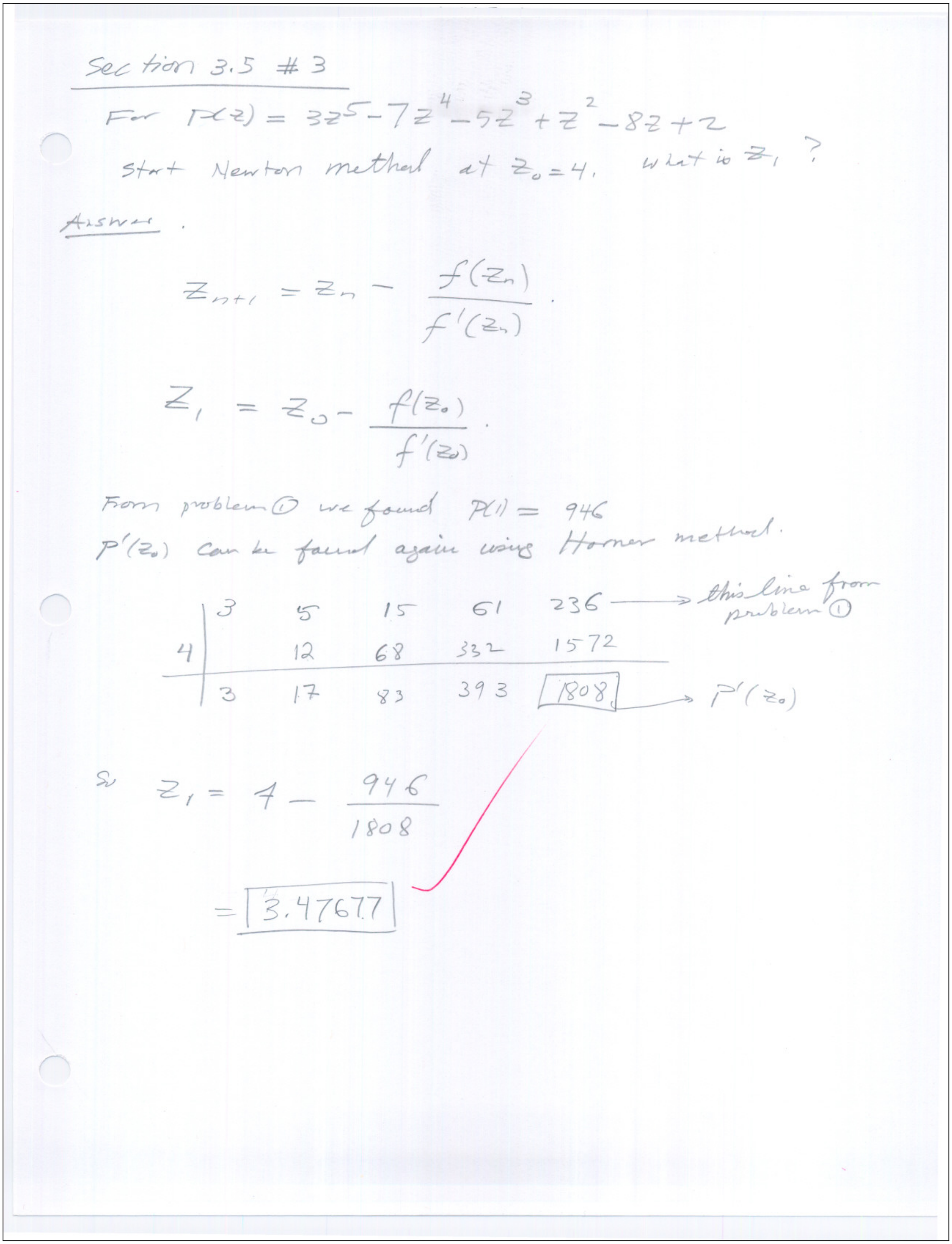
$$P'''(z) = 180z^2 - 168z - 30 \quad @ z_0=4 \Rightarrow \boxed{2178}$$

$$P''''(z) = 360z - 168 \quad @ z_0=4 \Rightarrow \boxed{1272}$$

$$P^{(5)}z = 360$$

so $$P(z) = 946 + (z-4)1808 + \frac{(z-4)^2 2378}{2!} + \frac{(z-4)^3 2178}{3!} + \frac{(z-4)^4 1272}{4!} + \frac{(z-4)^5 360}{5!}$$

$$\boxed{P(z) = 946 + 1808(z-4) + 1189(z-4)^2 + 363(z-4)^3 + 53(z-4)^4 + 3(z-4)^5}$$



Section 3.5 # 3

For $P(z) = 3z^5 - 7z^4 - 5z^3 + z^2 - 8z + 2$

start Newton method at $z_0 = 4$. what is z_1 ?

Answer.

$$z_{n+1} = z_n - \frac{f(z_n)}{f'(z_n)}.$$

$$z_1 = z_0 - \frac{f(z_0)}{f'(z_0)}.$$

From problem ① we found $P(1) = 946$

$P'(z_0)$ can be found again using Horner method.

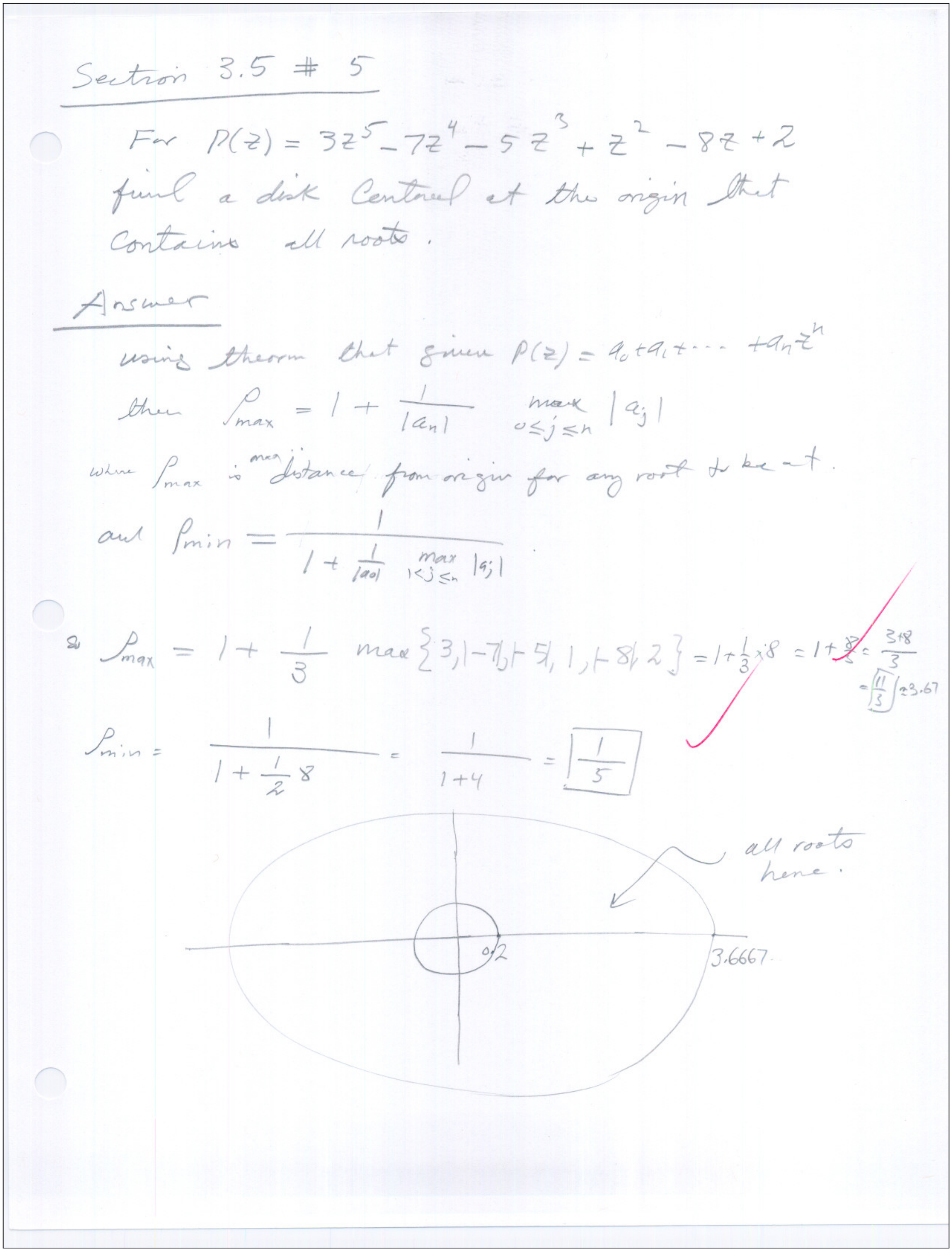
$$\begin{array}{c|ccccc}
 & 3 & 5 & 15 & 61 & 236 \\
4 & & 12 & 68 & 332 & 1572 \\
\hline
 & 3 & 17 & 83 & 393 & \boxed{1808}
\end{array}$$

236 → this line from problem ①

$\boxed{1808} \rightarrow P'(z_0)$

So $z_1 = 4 - \frac{946}{1808}$

$= \boxed{3.47677}$



Section 3.5 # 5

For $P(z) = 3z^5 - 7z^4 - 5z^3 + z^2 - 8z + 2$ find a disk Centered at the origin that Contains all roots.

Answer

using theorm that given $P(z) = a_0 + a_1 z + \cdots + a_n z^n$

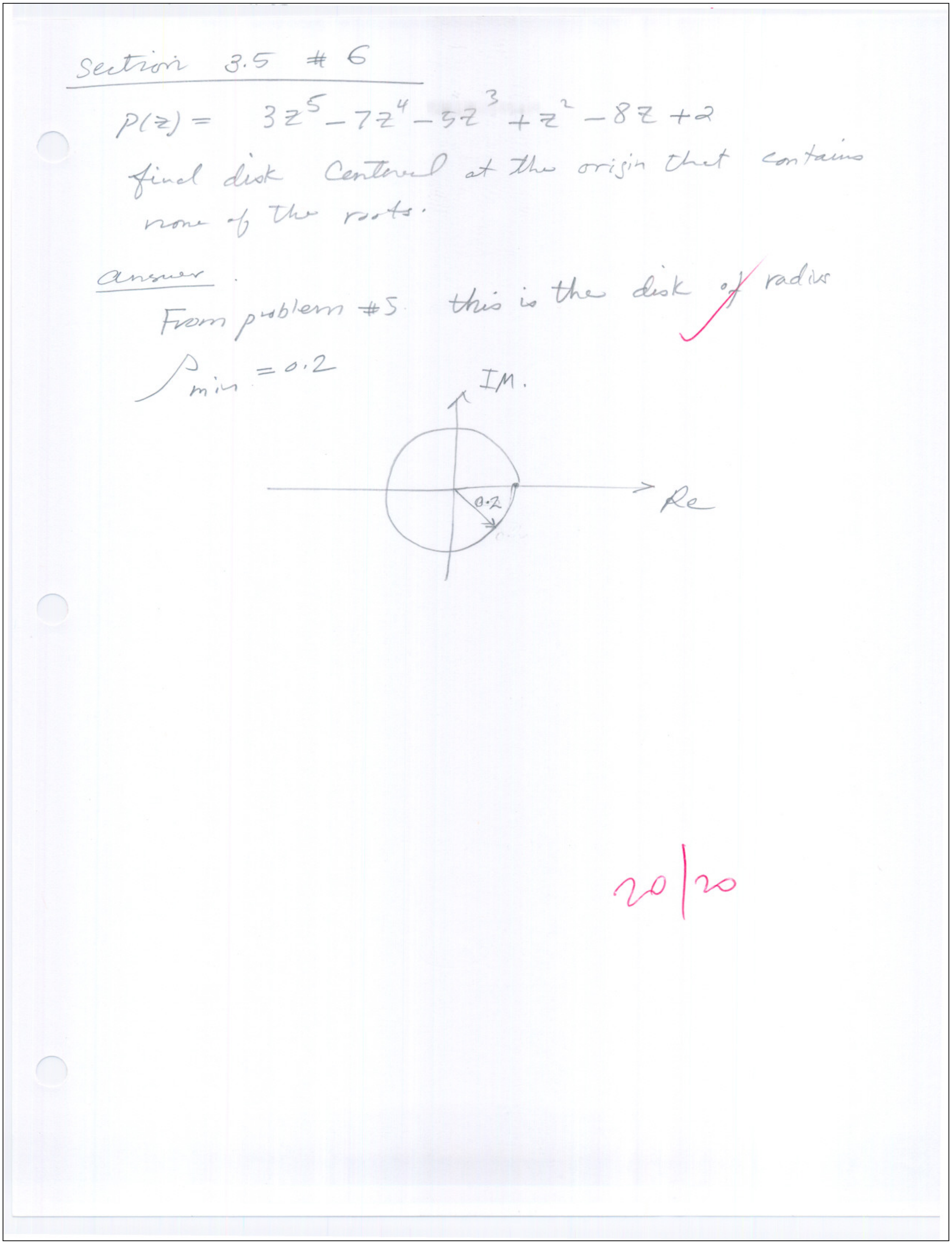
then $\rho_{max} = 1 + \frac{1}{|a_n|} \max_{0 \leq j \leq n} |a_j|$

where ρ_{max} is max distance from origin for any root to be at.

and $\rho_{min} = \frac{1}{1 + \frac{1}{|a_0|} \max_{1 < j \leq n} |a_j|}$.

so $\rho_{max} = 1 + \frac{1}{3} \max\{3, |-7|, |-5|, 1, |-8|, 2\} = 1 + \frac{1}{3} \times 8 = 1 + \frac{8}{3} = \frac{3+8}{3} = \boxed{\frac{11}{3}} \approx 3.67$ ✓

$\rho_{min} = \frac{1}{1 + \frac{1}{2} 8} = \frac{1}{1+4} = \boxed{\frac{1}{5}}$



Section 3.5 # 6

$p(z) = 3z^5 - 7z^4 - 5z^3 + z^2 - 8z + 2$

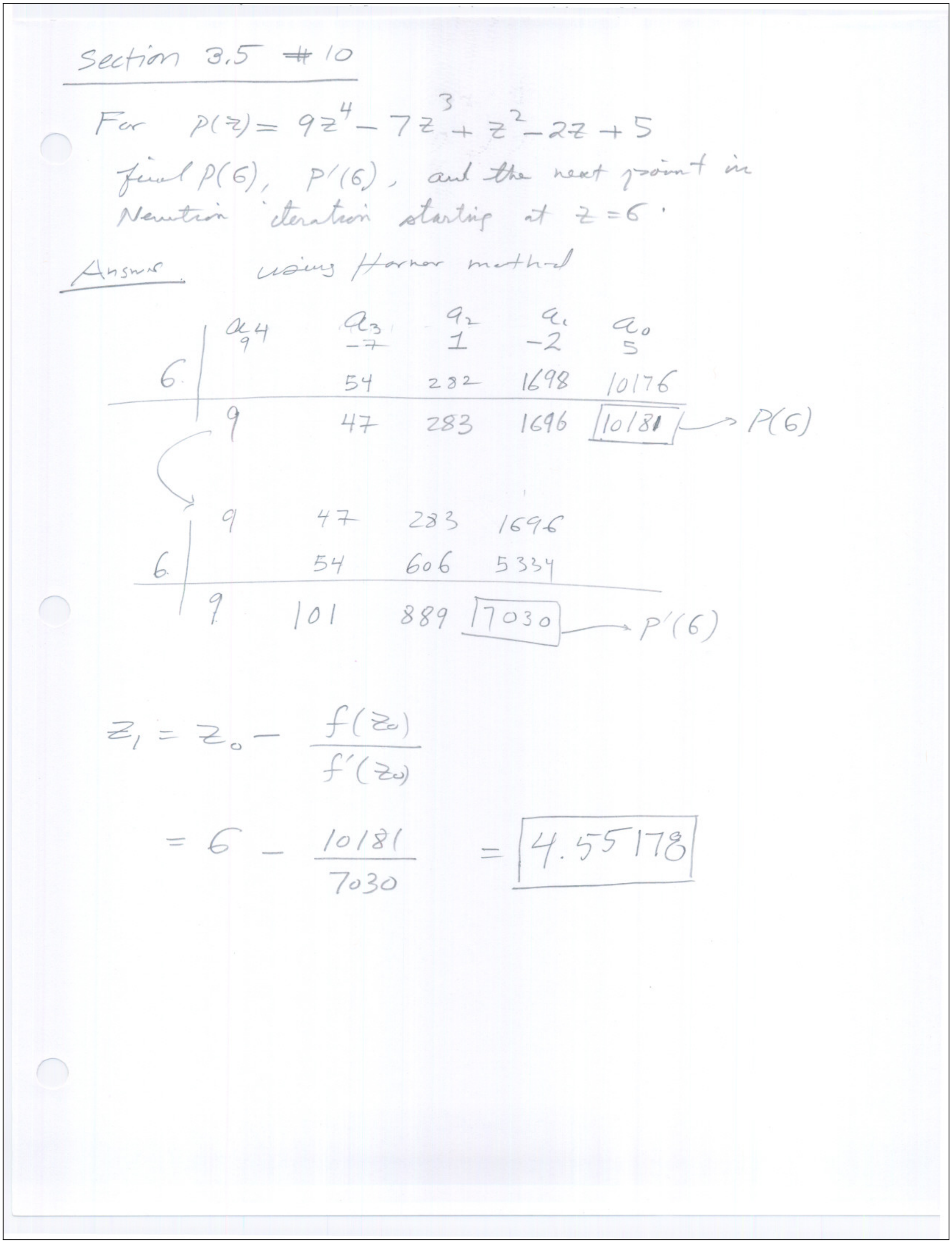
find disk Centered at the origin that contains none of the roots.

answer.

From problem #5. this is the disk of radius

$\rho_{min} = 0.2$

20|20



Section 3.5 #10

For $p(z) = 9z^4 - 7z^3 + z^2 - 2z + 5$

find $P(6)$, $P'(6)$, and the next point in Newton iteration starting at $z=6$.

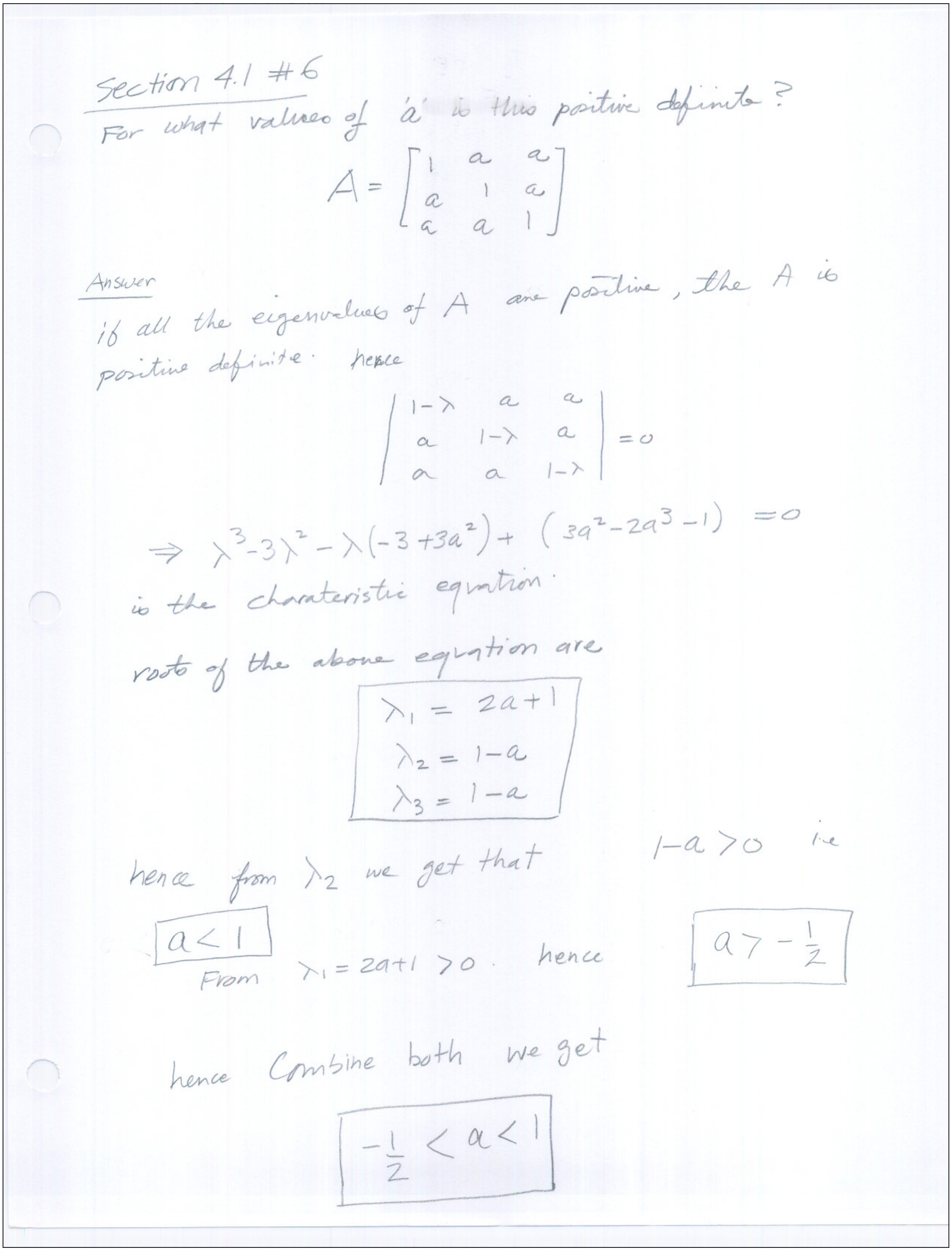
Answer using Horner method

$$\begin{array}{c|ccccc}
 & a_4 & a_3 & a_2 & a_1 & a_0 \\
 & 9 & -7 & 1 & -2 & 5 \\
6 & & 54 & 282 & 1698 & 10176 \\
\hline
 & 9 & 47 & 283 & 1696 & \boxed{10181} \longrightarrow P(6)
\end{array}$$

$$\begin{array}{c|cccc}
 & 9 & 47 & 283 & 1696 \\
6 & & 54 & 606 & 5334 \\
\hline
 & 9 & 101 & 889 & \boxed{7030} \longrightarrow P'(6)
\end{array}$$

$$z_1 = z_0 - \frac{f(z_0)}{f'(z_0)}$$

$$= 6 - \frac{10181}{7030} = \boxed{4.55178}$$



Section 4.1 #6

For what values of 'a' is this positive definite?

$$A=\begin{bmatrix}1 & a & a\\ a & 1 & a\\ a & a & 1\end{bmatrix}$$

Answer

if all the eigenvalues of A are positive, the A is positive definite. hence

$$\begin{vmatrix}1-\lambda & a & a\\ a & 1-\lambda & a\\ a & a & 1-\lambda\end{vmatrix}=0$$

$$\Rightarrow \lambda^3-3\lambda^2-\lambda(-3+3a^2)+(3a^2-2a^3-1)=0$$

is the charateristic equation.

roots of the above equation are

$$\boxed{\begin{aligned}\lambda_1&=2a+1\\ \lambda_2&=1-a\\ \lambda_3&=1-a\end{aligned}}$$

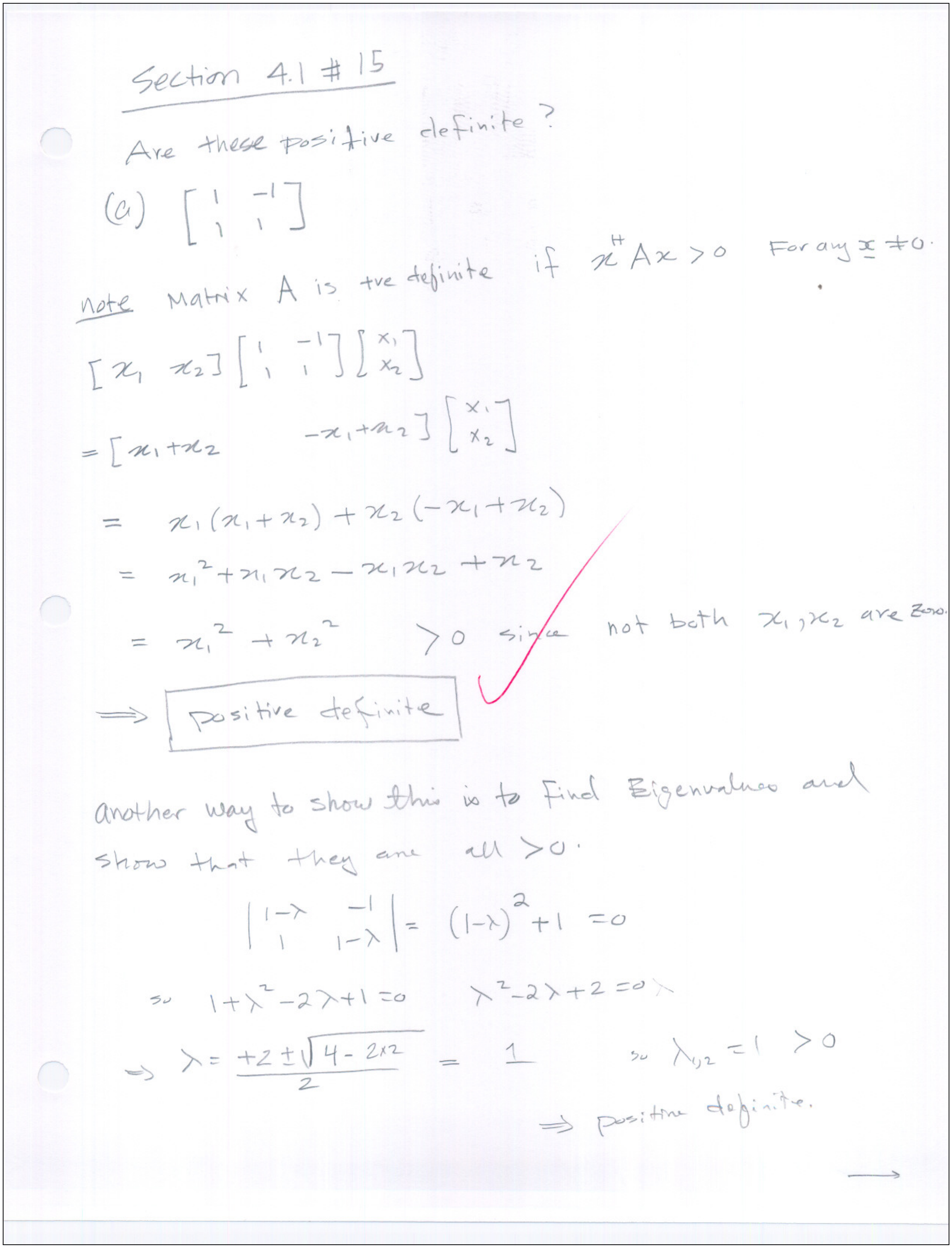
hence from λ_2 we get that $1-a>0$ i.e

$\boxed{a<1}$

From $\lambda_1=2a+1>0$. hence $\boxed{a>-\frac{1}{2}}$

hence Combine both we get

$$\boxed{-\frac{1}{2}<a<1}$$



Section 4.1 # 15

Are these positive definite?

(a) $\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$

note Matrix A is +ve definite if $x^H A x > 0$ For any $\underline{x} \neq 0$.

$[x_1 \quad x_2]\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= [x_1 + x_2 \qquad -x_1 + x_2]\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$= x_1(x_1 + x_2) + x_2(-x_1 + x_2)$

$= x_1^2 + x_1 x_2 - x_1 x_2 + x_2$

$= x_1^2 + x_2^2 \quad > 0$ since not both x_1, x_2 are zero.

$\Longrightarrow$ $\boxed{\text{positive definite}}$ ✓

another way to show this is to find Eigenvalues and show that they are all > 0.

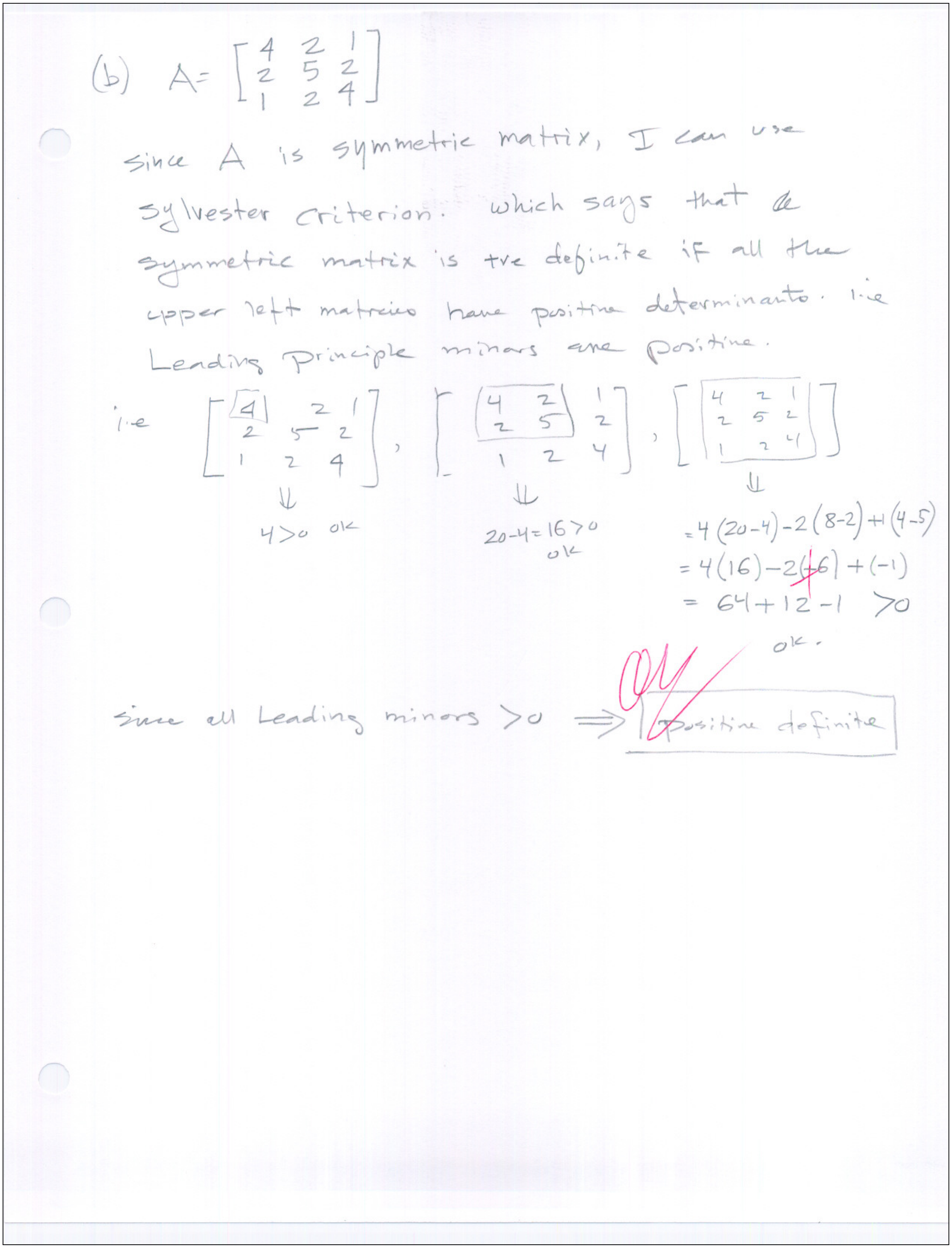
$$\begin{vmatrix} 1-\lambda & -1 \\ 1 & 1-\lambda \end{vmatrix} = (1-\lambda)^2 + 1 = 0$$

so $1 + \lambda^2 - 2\lambda + 1 = 0 \qquad \lambda^2 - 2\lambda + 2 = 0$

$\Rightarrow \lambda = \dfrac{+2 \pm \sqrt{4 - 2 \times 2}}{2} = 1$ so $\lambda_{1,2} = 1 > 0$

$\Rightarrow$ positive definite.

→



(b) $A = \begin{bmatrix} 4 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix}$

since A is symmetric matrix, I can use sylvester criterion. which says that ~~a~~ symmetric matrix is +ve definite if all the upper left matrices have positive determinants. i.e Leading principle minors are positive.

i.e $\begin{bmatrix} \boxed{4} & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{bmatrix}$, $\begin{bmatrix} \boxed{\begin{matrix} 4 & 2 \\ 2 & 5 \end{matrix}} & \begin{matrix} 1 \\ 2 \end{matrix} \\ \begin{matrix} 1 & 2 \end{matrix} & 4 \end{bmatrix}$, $\begin{bmatrix} \boxed{\begin{matrix} 4 & 2 & 1 \\ 2 & 5 & 2 \\ 1 & 2 & 4 \end{matrix}} \end{bmatrix}$

$\Downarrow$ $4 > 0$ ok

$\Downarrow$ $20 - 4 = 16 > 0$ ok

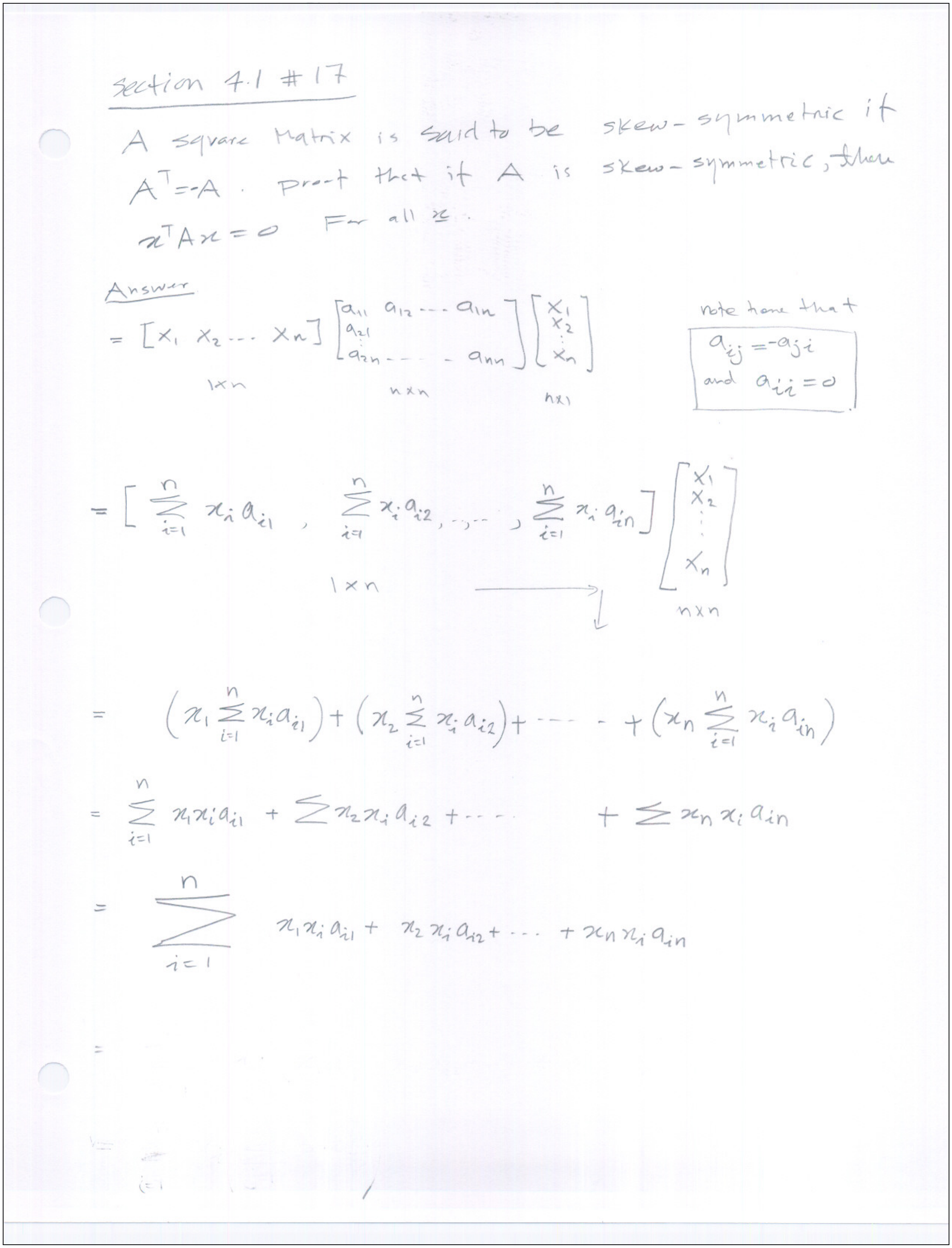
$\Downarrow$

$$= 4(20-4) - 2(8-2) + 1(4-5)$$
$$= 4(16) - 2(+6) + (-1)$$
$$= 64 + 12 - 1 > 0$$

ok.

Ok

since all Leading minors $> 0 \implies$ $\boxed{\text{Positive definite}}$



section 4.1 #17

A square Matrix is said to be skew-symmetric if $A^T=-A$. proof that if A is skew-symmetric, then $x^TAx=0$ For all x.

Answer

$$= \underset{1\times n}{\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}} \underset{n\times n}{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & & & \\ a_{2n} & \cdots & & a_{nn} \end{bmatrix}} \underset{n\times 1}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}$$

note here that $\boxed{\begin{array}{l} a_{ij} = -a_{ji} \\ \text{and } a_{ii} = 0 \end{array}}$

$$= \underset{1\times n}{\left[\sum_{i=1}^{n} x_i a_{i1} \;,\; \sum_{i=1}^{n} x_i a_{i2} , \ldots , \sum_{i=1}^{n} x_i a_{in} \right]} \xrightarrow{\qquad} \underset{n\times n}{\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}}$$

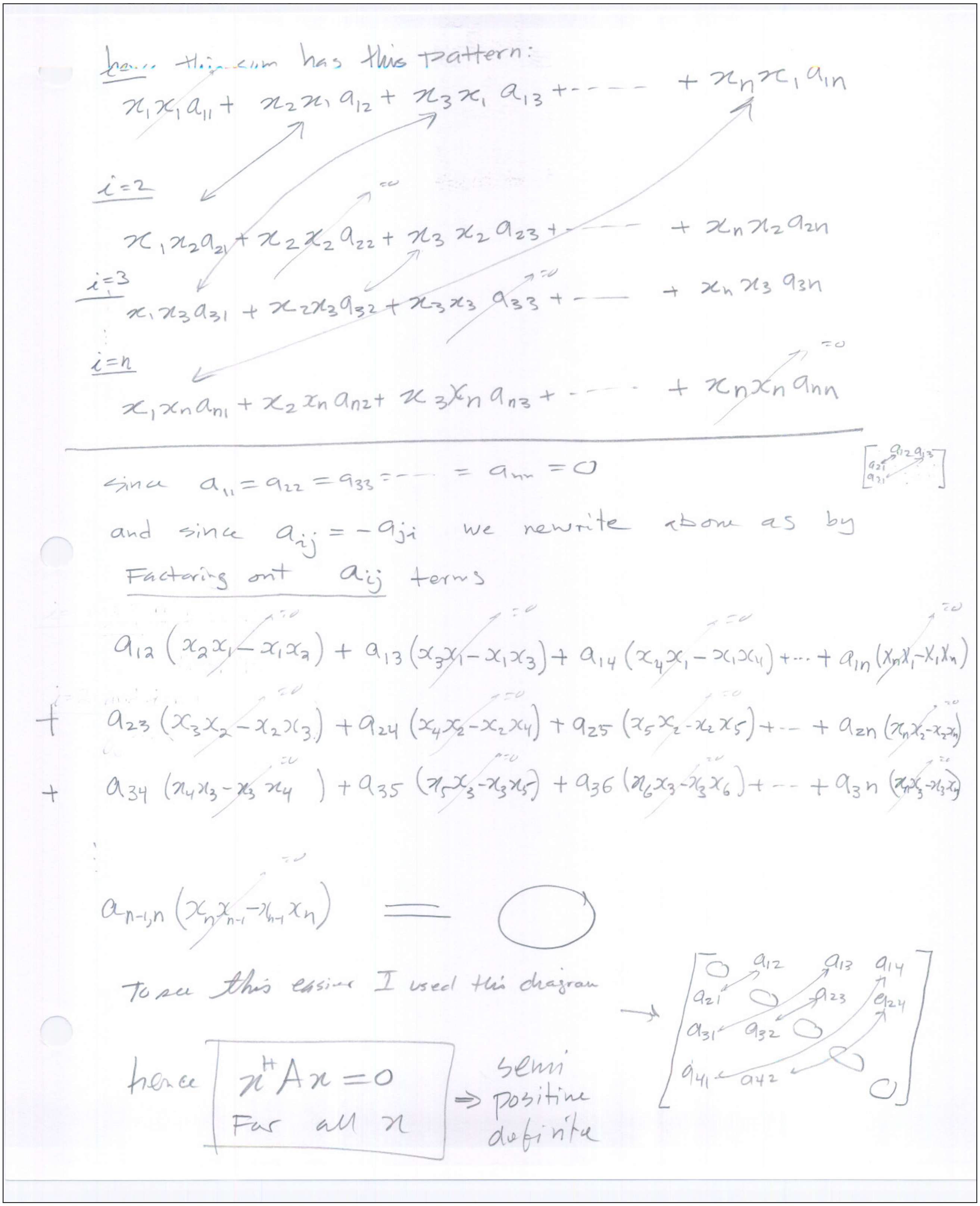
$$\downarrow$$

$$= \left(x_1 \sum_{i=1}^{n} x_i a_{i1} \right) + \left(x_2 \sum_{i=1}^{n} x_i a_{i2} \right) + \cdots + \left(x_n \sum_{i=1}^{n} x_i a_{in} \right)$$

$$= \sum_{i=1}^{n} x_1 x_i a_{i1} + \sum x_2 x_i a_{i2} + \cdots + \sum x_n x_i a_{in}$$

$$= \sum_{i=1}^{n} x_1 x_i a_{i1} + x_2 x_i a_{i2} + \cdots + x_n x_i a_{in}$$

$$=$$



hence this sum has this pattern:

$$\cancel{x_1x_1a_{11}} + x_2x_1a_{12} + x_3x_1a_{13} + \cdots + x_nx_1a_{1n}$$

$i=2$

$$x_1x_2a_{21} + \overset{=0}{\cancel{x_2x_2a_{22}}} + x_3x_2a_{23} + \cdots + x_nx_2a_{2n}$$

$i=3$

$$x_1x_3a_{31} + x_2x_3a_{32} + \overset{=0}{\cancel{x_3x_3a_{33}}} + \cdots + x_nx_3a_{3n}$$

$\vdots$

$i=n$

$$x_1x_na_{n1} + x_2x_na_{n2} + x_3x_na_{n3} + \cdots + \overset{=0}{\cancel{x_nx_na_{nn}}}$$

$$\begin{bmatrix} & a_{12} & a_{13} \\ a_{21} & & \\ a_{31} & & \end{bmatrix}$$

since $a_{11}=a_{22}=a_{33}=\cdots=a_{nn}=0$

and since $a_{ij}=-a_{ji}$ we rewrite above as by Factoring out a_{ij} terms

$$a_{12}\overset{=0}{\cancel{(x_2x_1-x_1x_2)}} + a_{13}\overset{=0}{\cancel{(x_3x_1-x_1x_3)}} + a_{14}\overset{=0}{\cancel{(x_4x_1-x_1x_4)}} + \cdots + a_{1n}\overset{=0}{\cancel{(x_nx_1-x_1x_n)}}$$

$$+\ a_{23}\overset{=0}{\cancel{(x_3x_2-x_2x_3)}} + a_{24}\overset{=0}{\cancel{(x_4x_2-x_2x_4)}} + a_{25}\overset{=0}{\cancel{(x_5x_2-x_2x_5)}} + \cdots + a_{2n}\overset{=0}{\cancel{(x_nx_2-x_2x_n)}}$$

$$+\ a_{34}\overset{=0}{\cancel{(x_4x_3-x_3x_4)}} + a_{35}\overset{=0}{\cancel{(x_5x_3-x_3x_5)}} + a_{36}\overset{=0}{\cancel{(x_6x_3-x_3x_6)}} + \cdots + a_{3n}\overset{=0}{\cancel{(x_nx_3-x_3x_n)}}$$

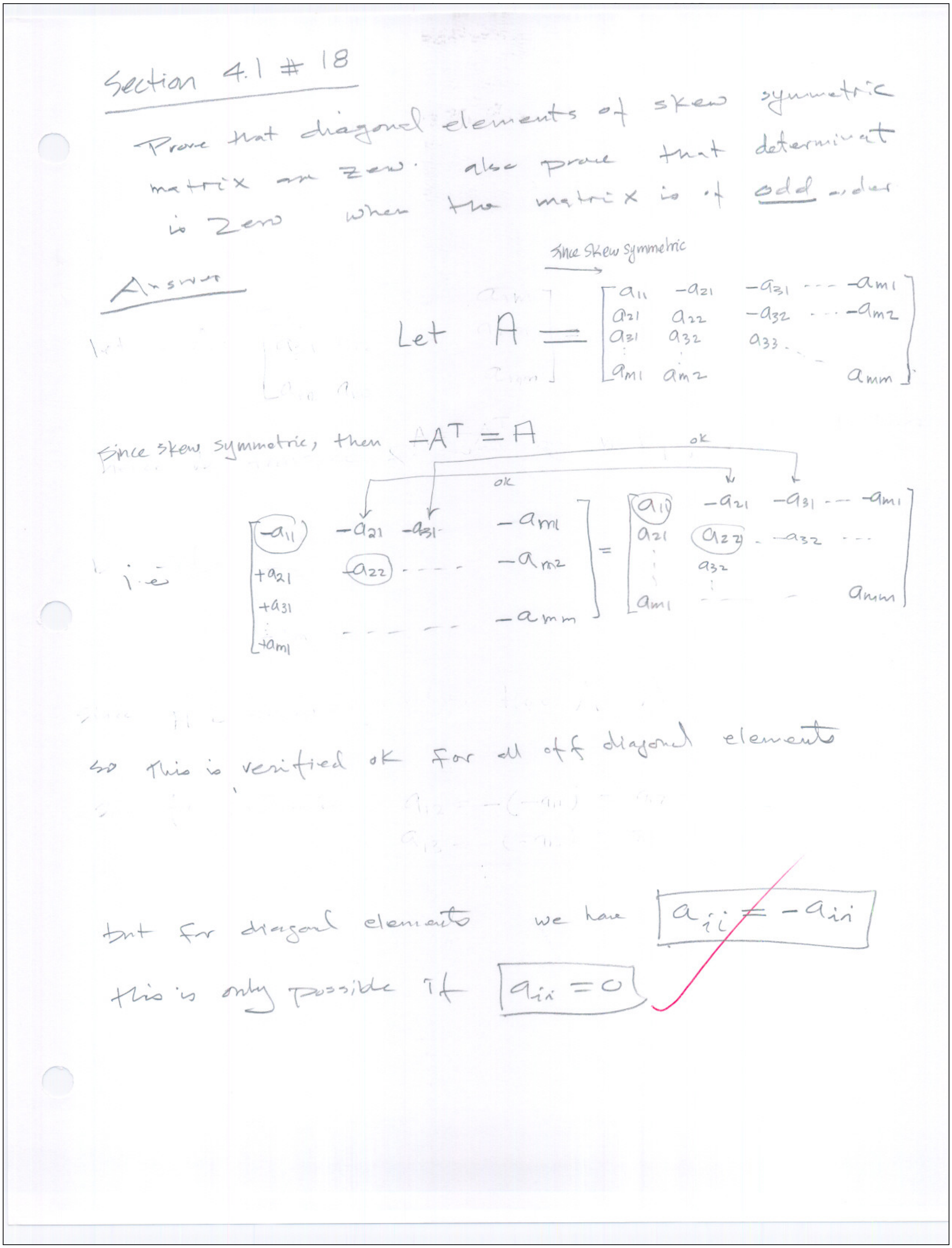
$\vdots$

$$a_{n-1,n}\overset{=0}{\cancel{(x_nx_{n-1}-x_{n-1}x_n)}} = 0$$

To see this easier I used this diagram →

$$\begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ a_{21} & 0 & a_{23} & a_{24} \\ a_{31} & a_{32} & 0 & \\ a_{41} & a_{42} & & 0 \end{bmatrix}$$

hence $\boxed{x^H A x = 0 \text{ For all } x}$ ⇒ semi positive definite



Section 4.1 # 18

Prove that diagonal elements of skew symmetric matrix are zero. also prove that determinat is zero when the matrix is of odd order

Answer

Since skew symmetric →

$$\text{Let } A = \begin{bmatrix} a_{11} & -a_{21} & -a_{31} & \cdots & -a_{m1} \\ a_{21} & a_{22} & -a_{32} & \cdots & -a_{m2} \\ a_{31} & a_{32} & a_{33} & \cdots & \\ \vdots & \vdots & & & \\ a_{m1} & a_{m2} & & & a_{mm} \end{bmatrix}$$

Since skew symmetric, then $-A^T = A$

ok

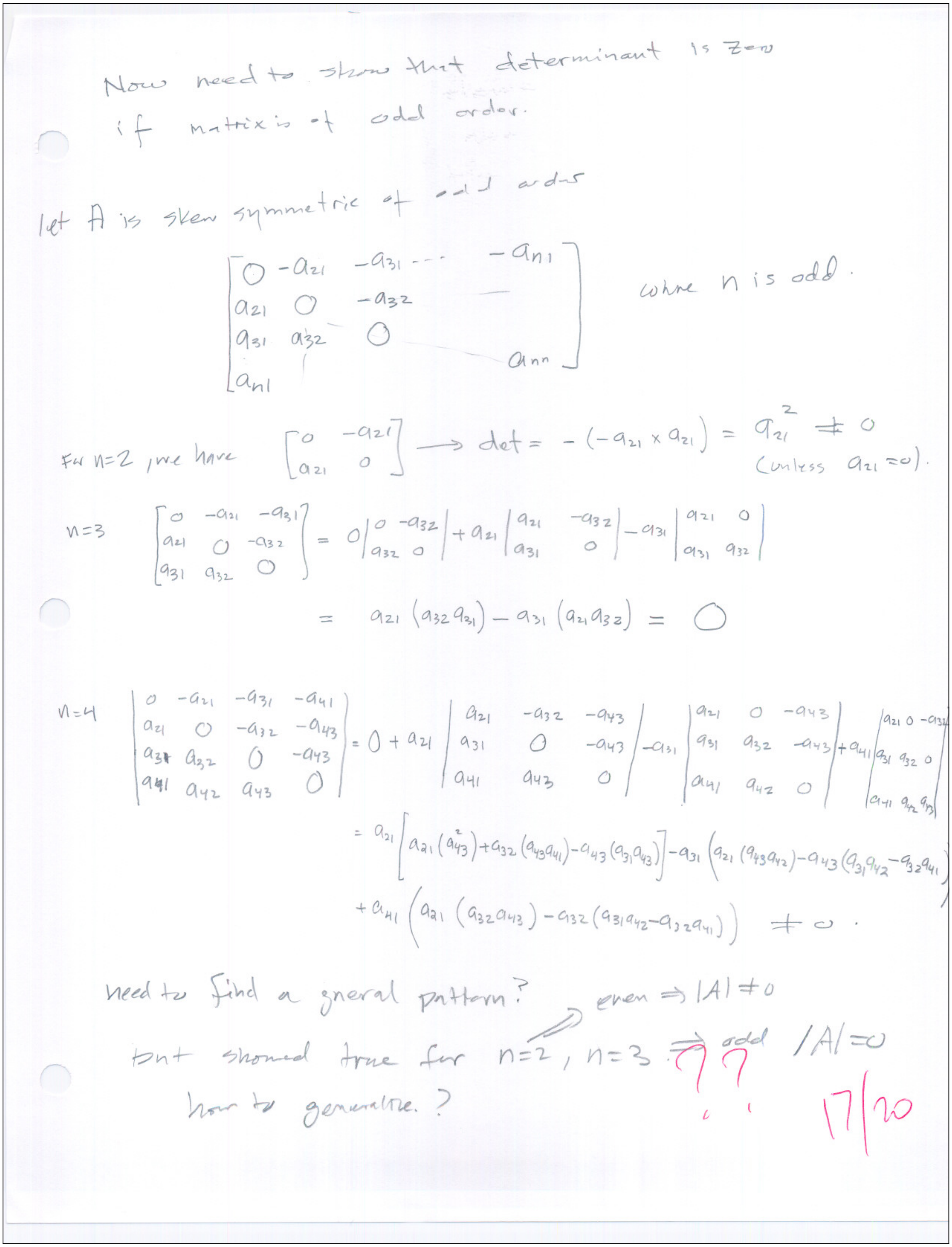
ok

$$\text{i.e.} \quad \begin{bmatrix} -a_{11} & -a_{21} & -a_{31} & & -a_{m1} \\ +a_{21} & -a_{22} & \cdots & & -a_{m2} \\ +a_{31} & & & & \\ \vdots & & & & \\ +a_{m1} & & \cdots & & -a_{mm} \end{bmatrix} = \begin{bmatrix} a_{11} & -a_{21} & -a_{31} & \cdots & -a_{m1} \\ a_{21} & a_{22} & -a_{32} & \cdots & \\ \vdots & a_{32} & & & \\ \vdots & \vdots & & & \\ a_{m1} & \cdots & & & a_{mm} \end{bmatrix}$$

so this is verified ok for all off diagonal elements

but for diagonal elements we have $\boxed{a_{ii} = -a_{ii}}$

this is only possible if $\boxed{a_{ii} = 0}$



Now need to show that determinant is zero

if matrix is of odd order.

let A is skew symmetric of odd order

$$\begin{bmatrix} 0 & -a_{21} & -a_{31} & \cdots & -a_{n1} \\ a_{21} & 0 & -a_{32} & & \\ a_{31} & a_{32} & 0 & & \\ & & & & a_{nn} \\ a_{n1} & & & & \end{bmatrix}$$ where n is odd.

For $n=2$, we have $\begin{bmatrix} 0 & -a_{21} \\ a_{21} & 0 \end{bmatrix} \longrightarrow \det = -(-a_{21} \times a_{21}) = a_{21}^2 \neq 0$ (unless $a_{21}=0$).

$n=3$ $$\begin{bmatrix} 0 & -a_{21} & -a_{31} \\ a_{21} & 0 & -a_{32} \\ a_{31} & a_{32} & 0 \end{bmatrix} = 0\begin{vmatrix} 0 & -a_{32} \\ a_{32} & 0 \end{vmatrix} + a_{21}\begin{vmatrix} a_{21} & -a_{32} \\ a_{31} & 0 \end{vmatrix} - a_{31}\begin{vmatrix} a_{21} & 0 \\ a_{31} & a_{32} \end{vmatrix}$$

$$= a_{21}(a_{32}a_{31}) - a_{31}(a_{21}a_{32}) = 0$$

$n=4$ $$\begin{vmatrix} 0 & -a_{21} & -a_{31} & -a_{41} \\ a_{21} & 0 & -a_{32} & -a_{43} \\ a_{31} & a_{32} & 0 & -a_{43} \\ a_{41} & a_{42} & a_{43} & 0 \end{vmatrix} = 0 + a_{21}\begin{vmatrix} a_{21} & -a_{32} & -a_{43} \\ a_{31} & 0 & -a_{43} \\ a_{41} & a_{43} & 0 \end{vmatrix} - a_{31}\begin{vmatrix} a_{21} & 0 & -a_{43} \\ a_{31} & a_{32} & -a_{43} \\ a_{41} & a_{42} & 0 \end{vmatrix} + a_{41}\begin{vmatrix} a_{21} & 0 & -a_{32} \\ a_{31} & a_{32} & 0 \\ a_{41} & a_{42} & a_{43} \end{vmatrix}$$

$$= a_{21}\left[a_{21}(a_{43}^2) + a_{32}(a_{43}a_{41}) - a_{43}(a_{31}a_{43})\right] - a_{31}\left(a_{21}(a_{43}a_{42}) - a_{43}(a_{31}a_{42} - a_{32}a_{41})\right)$$

$$+ a_{41}\left(a_{21}(a_{32}a_{43}) - a_{32}(a_{31}a_{42} - a_{32}a_{41})\right) \neq 0.$$

need to find a general pattern? even $\Rightarrow |A| \neq 0$

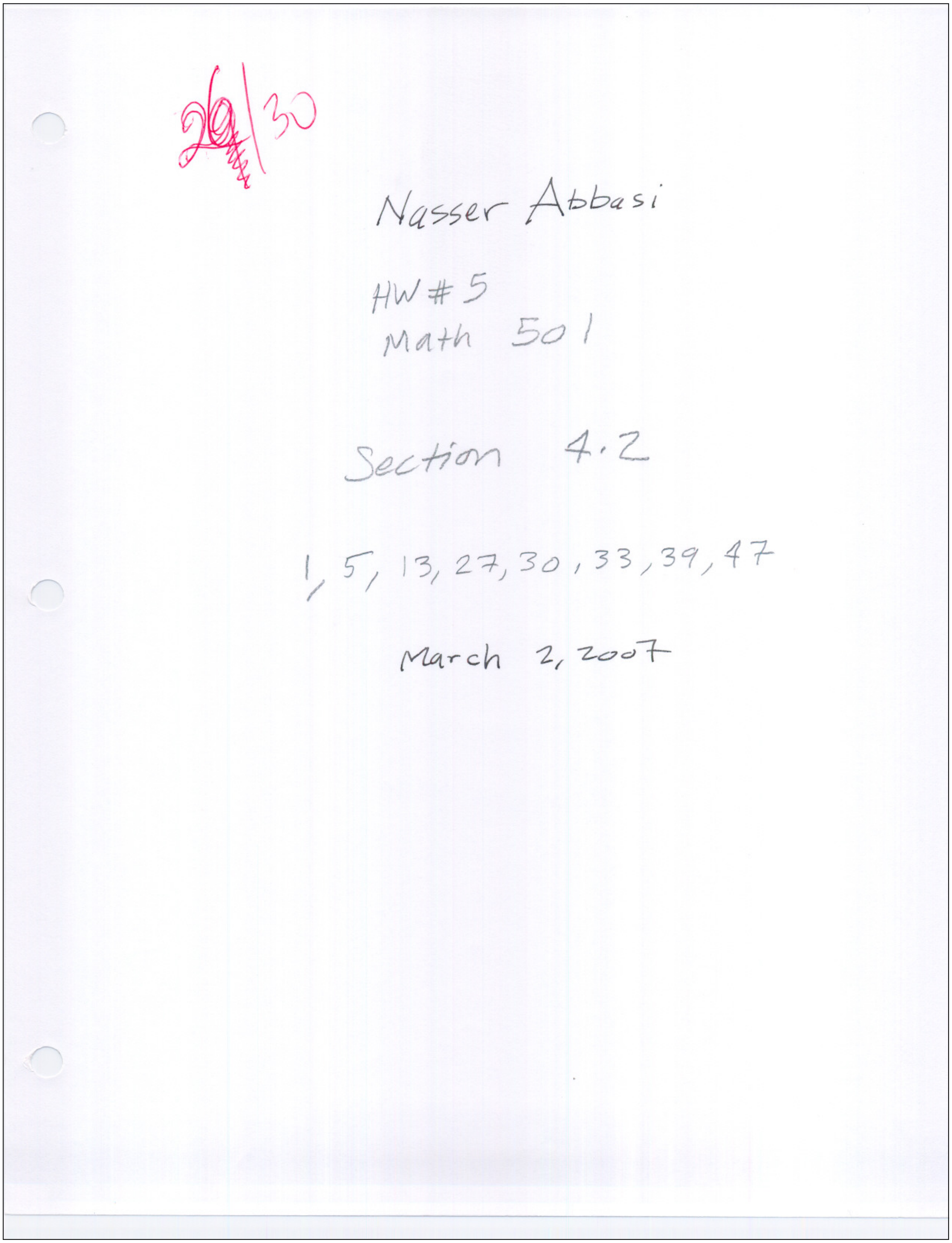
but showed true for $n=2$, $n=3 \Rightarrow$ odd $|A|=0$

how to generalize. ?

??

17/20

4.6 HW 5



26/30

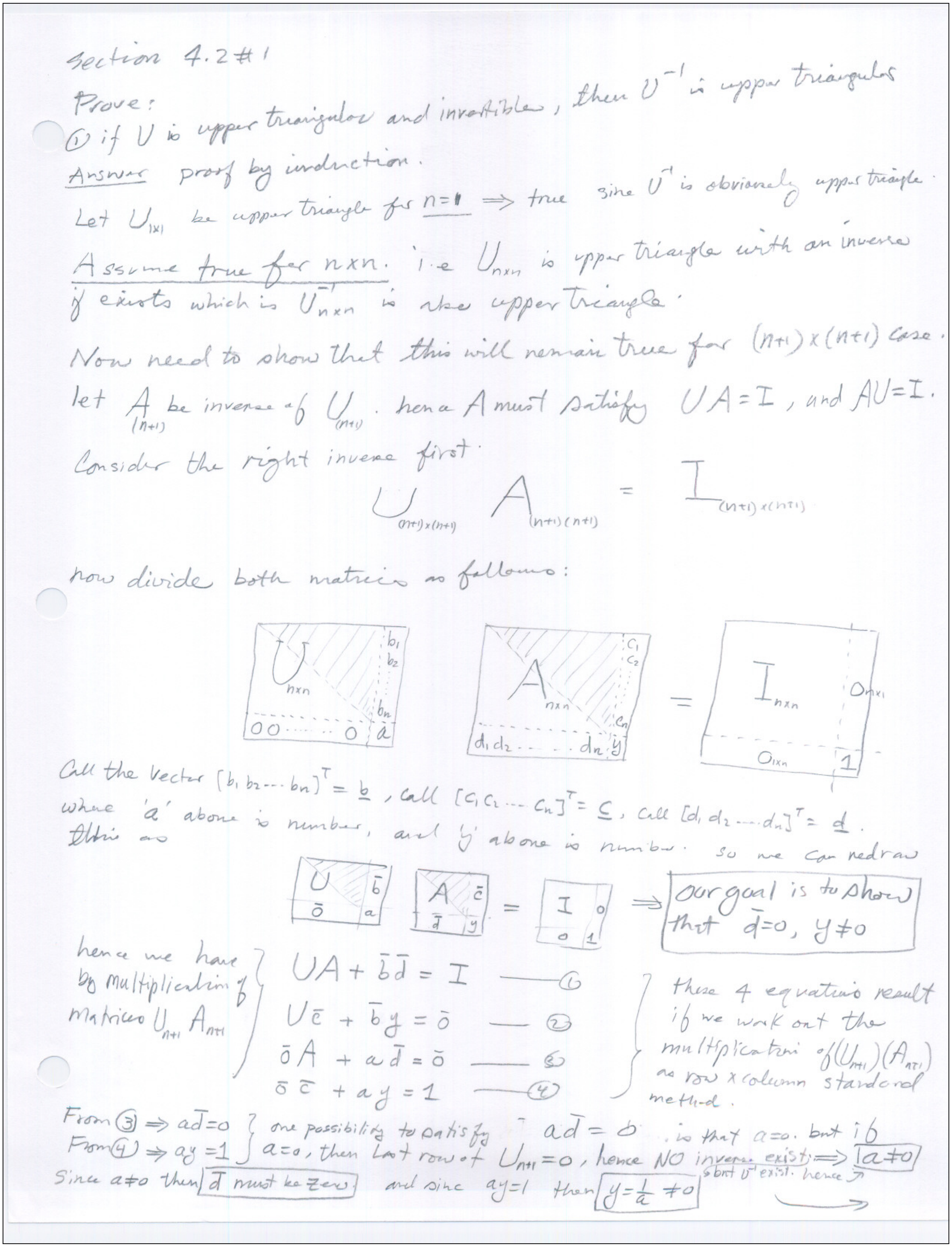
Nasser Abbasi

HW#5
Math 501

Section 4.2

1, 5, 13, 27, 30, 33, 39, 47

March 2, 2007



section 4.2#1

Prove:

① if U is upper triangular and invertible, then U^{-1} is upper triangular

Answer proof by induction.

Let $U_{1\times 1}$ be upper triangle for $n=1$ $\Rightarrow$ true since U^{-1} is obviously upper triangle.

Assume true for $n\times n$. i.e $U_{n\times n}$ is upper triangle with an inverse it exists which is $U^{-1}_{n\times n}$ is also upper triangle.

Now need to show that this will remain true for $(n+1)\times(n+1)$ case.

let $A_{(n+1)}$ be inverse of $U_{(n+1)}$. hence A must satisfy $UA=I$, and $AU=I$.

Consider the right inverse first.

$$U_{(n+1)\times(n+1)} \; A_{(n+1)(n+1)} = I_{(n+1)\times(n+1)}$$

now divide both matrices as follows:

$$\left[\begin{array}{cccc|c} & & & & b_1 \\ & U_{n\times n} & & & b_2 \\ & & & & \vdots \\ & & & & b_n \\ \hline 0 & 0 & \cdots & 0 & a \end{array}\right] \left[\begin{array}{cccc|c} & & & & c_1 \\ & A_{n\times n} & & & c_2 \\ & & & & \vdots \\ & & & & c_n \\ \hline d_1 & d_2 & \cdots & d_n & y \end{array}\right] = \left[\begin{array}{c|c} I_{n\times n} & 0_{n\times 1} \\ \hline 0_{1\times n} & 1 \end{array}\right]$$

Call the vector $[b_1\, b_2 \cdots b_n]^T = \underline{b}$, call $[c_1\, c_2 \cdots c_n]^T = \underline{c}$, call $[d_1\, d_2 \cdots d_n]^T = \underline{d}$.

where 'a' above is number, and 'y' above is number. so we can redraw this as

$$\left[\begin{array}{c|c} U & \bar{b} \\ \hline \bar{o} & a \end{array}\right] \left[\begin{array}{c|c} A & \bar{c} \\ \hline \bar{d} & y \end{array}\right] = \left[\begin{array}{c|c} I & 0 \\ \hline 0 & 1 \end{array}\right] \Rightarrow \boxed{\text{our goal is to show that } \bar{d}=0,\ y\neq 0}$$

hence we have by multiplication of matrices U_{n+1} A_{n+1}

$$UA + \bar{b}\bar{d} = I \text{ —— ①}$$

$$U\bar{c} + \bar{b}y = \bar{o} \text{ —— ②}$$

$$\bar{o}A + a\bar{d} = \bar{o} \text{ —— ③}$$

$$\bar{o}\,\bar{c} + ay = 1 \text{ —— ④}$$

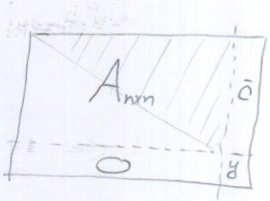
these 4 equations result if we work out the multiplication of $(U_{n+1})(A_{n+1})$ as row x column standard method.

From ③ $\Rightarrow a\bar{d}=0$
From ④ $\Rightarrow ay=1$

one possibility to satisfy $a\bar{d}=0$ is that $a=0$. but if $a=0$, then last row of $U_{n+1}=0$, hence NO inverse exist, but U^{-1} exist. hence $\Rightarrow \boxed{a\neq 0}$

Since $a\neq 0$ then $\boxed{\bar{d} \text{ must be zero}}$ and since $ay=1$ then $\boxed{y=\frac{1}{a}\neq 0}$ $\longrightarrow$

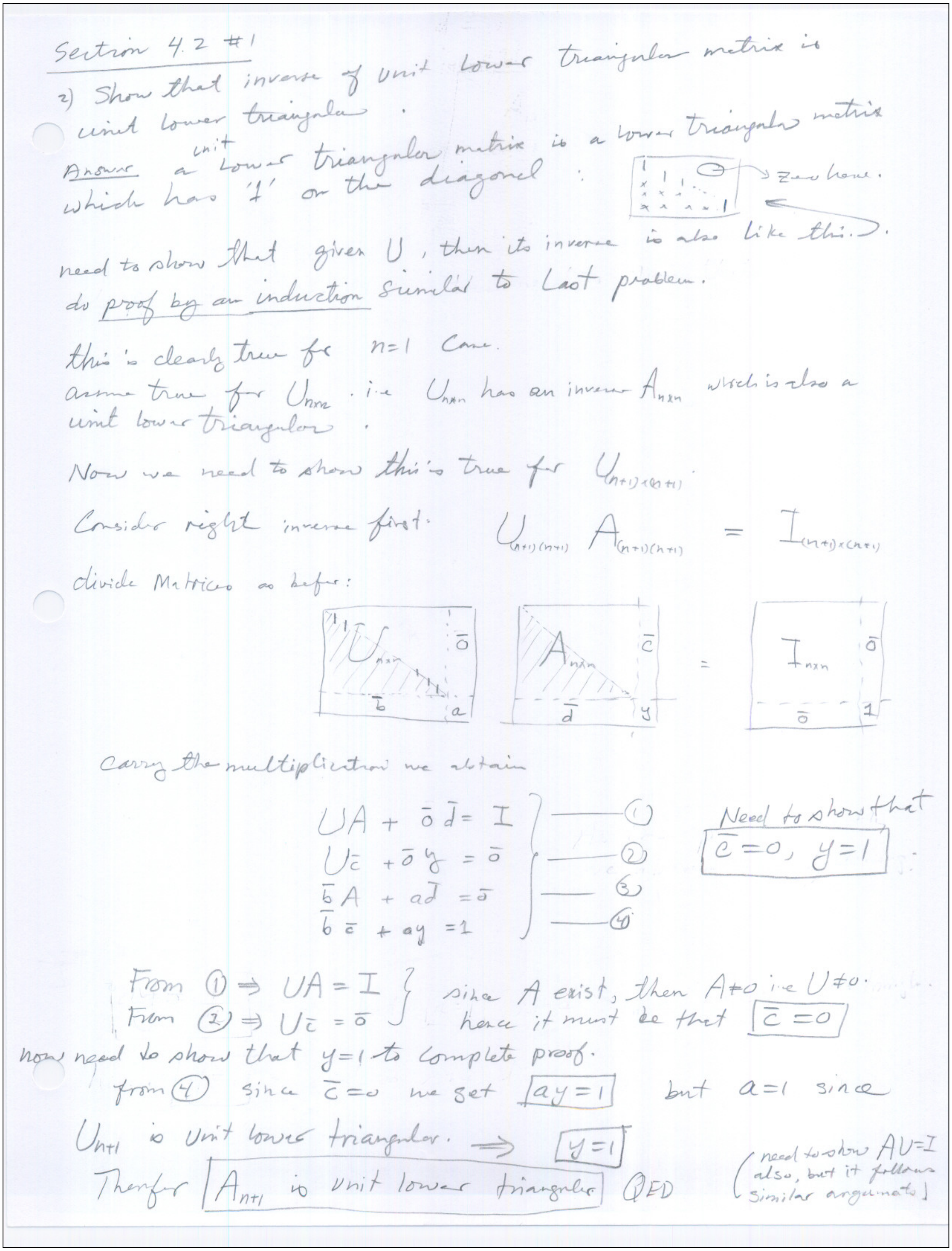
therefor we showed that $\bar{d}=0$ and $y\neq 0$. hence A_{n+1} is



but since we assumed that $A_{n\times n}$ is the inverse of $U_{n\times n}$ and is upper trianglur, then we have shown that $A_{(n+1)\times(n+1)}$ is also upper triangulr. and sine $U_{n\times 1} A_{n+1} = I_{n+1}$, its is the inverse of U_{n+1}.

we need to also show this for Left inverse. $A_{n\times 1} U_{n\times 1} = I_{n\times 1}$ but the argumet will follow the same lines.

QED.



Section 4.2 #1

2) Show that inverse of unit lower triangular matrix is unit lower triangular.

Answer a unit lower triangular matrix is a lower triangular matrix which has '1' on the diagonal:

$$\begin{bmatrix} 1 & & \\ x & 1 & 1 & 0 \\ x & x & \ddots & \\ x & x & x & 1 \end{bmatrix} \rightarrow \text{Zero here.}$$

need to show that given U, then its inverse is also like this.

do proof by an induction similar to Last problem.

this is clearly true for $n=1$ Case.

assume true for $U_{n\times n}$. i.e $U_{n\times n}$ has an inverse $A_{n\times n}$ which is also a unit lower triangular.

Now we need to show this is true for $U_{(n+1)\times(n+1)}$.

Consider right inverse first: $U_{(n+1)(n+1)}\ A_{(n+1)(n+1)} = I_{(n+1)\times(n+1)}$

divide Matrices as before:

$$\begin{bmatrix} U_{n\times n} & \bar{0} \\ \bar{b} & a \end{bmatrix} \begin{bmatrix} A_{n\times n} & \bar{c} \\ \bar{d} & y \end{bmatrix} = \begin{bmatrix} I_{n\times n} & \bar{0} \\ \bar{0} & 1 \end{bmatrix}$$

Carry the multiplication we obtain

$$\left.\begin{aligned} UA + \bar{0}\bar{d} &= I \\ U\bar{c} + \bar{0}y &= \bar{0} \\ \bar{b}A + a\bar{d} &= \bar{0} \\ \bar{b}\bar{c} + ay &= 1 \end{aligned}\right\} \begin{aligned} &\text{———} \textcircled{1} \\ &\text{———} \textcircled{2} \\ &\text{———} \textcircled{3} \\ &\text{———} \textcircled{4} \end{aligned} \qquad \boxed{\text{Need to show that } \bar{c}=0,\ y=1}$$

From ① ⇒ $UA = I$

From ② ⇒ $U\bar{c} = \bar{0}$

} since A exist, then $A\neq 0$ i.e $U\neq 0$. hence it must be that $\boxed{\bar{c}=0}$

now need to show that $y=1$ to complete proof.

from ④ since $\bar{c}=0$ we get $\boxed{ay=1}$ but $a=1$ since U_{n+1} is unit lower triangular. $\Longrightarrow \boxed{y=1}$

Therefor $\boxed{A_{n+1} \text{ is unit lower triangular}}$ QED

(need to show $AV=I$ also, but it follows similar argument)

section 4.2 # 1

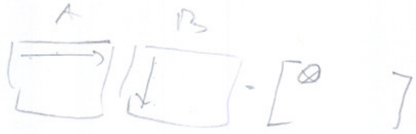
(c) show that product of 2 upper/lower triangular matrices is upper//lower triangular.

use definition of matrix multiplication:

$$\underset{n\times k}{C} = \underset{n\times m}{A}\ \underset{m\times k}{B}$$

where $C_{ij} = \sum_{q=1}^{m} A_{iq} B_{qj}$

row of A * column of B



for example $C_{11} = \sum_{q=1}^{m} A_{1q} B_{q1} = A_{11}B_{11} + A_{12}B_{21} + A_{13}B_{31}$

and $C_{12} = \sum_{q=1}^{m} A_{1q} B_{q2} = A_{11}B_{12} + A_{12}B_{22} + A_{13}B_{32}$

ok. verified.

now that matrix multiplication is defined, consider the first case: product of 2 upper triangular matrices:

since upper triangular, then

$A_{iq} = 0$ if $i > q$. i.e row number is > column number.

and $B_{qj} = 0$ if $q > j$

but $C_{ij} = \sum_{q=1}^{m} A_{iq} B_{qj}$.

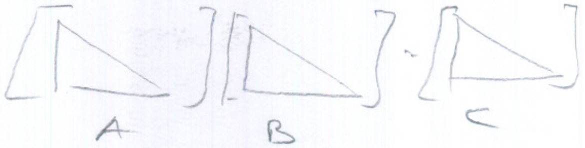
Then when $i > q$ or $q > j$ we have $A_{iq}B_{qj} = 0$.

Therefor when $i > q > j$ we have $A_{iq}B_{qj} = 0$

Therfor when $\boxed{i > j}$ we have $A_{iq}B_{qj} = 0$

$\Longrightarrow$ C is upper triangular as well. QED →

similar argument for the 2 lower triangular matrices.



here $A_{iq} = 0$ if $i<q$.

and $B_{qj} = 0$ if $q<j$

hence $C_{ij} = 0$ if $i<q<j$

or $C_{ij} = 0$ if $i<j$

but this means C is lower triangular.

section 4.2 #5

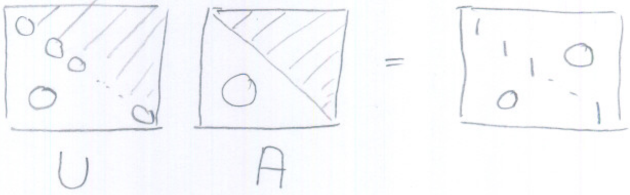
Proof that an upper or lower triangular Matrix is nonsingular iff its diagonal elements are all ≠0.

proof.

⟹ direction: show that if U is invertible then diagonal elements are all nonzero.

by contradiction: assume that U is upper triangle, and it has all its diagonal elements $=0$ and it has an Inverse, say A. then we have $UA=I$

i.e



Note I have used the fact that A is upper triangular. because we proved in problem 4.2 # 1 that an upper triangle will have an inverse which is also an upper triangle. (if inverse exist).

OK now we continue. Carry out the matrix multiplication, we obtain for first row of U:

$$u_{11}a_{11}+u_{12}a_{21}+\cdots+u_{1n}a_{n1}=\overset{I(1,1)}{1} \quad \text{——} \; ①$$

but $a_{21}=a_{31}=\cdots=a_{n1}=0$ since A is upper triangle.

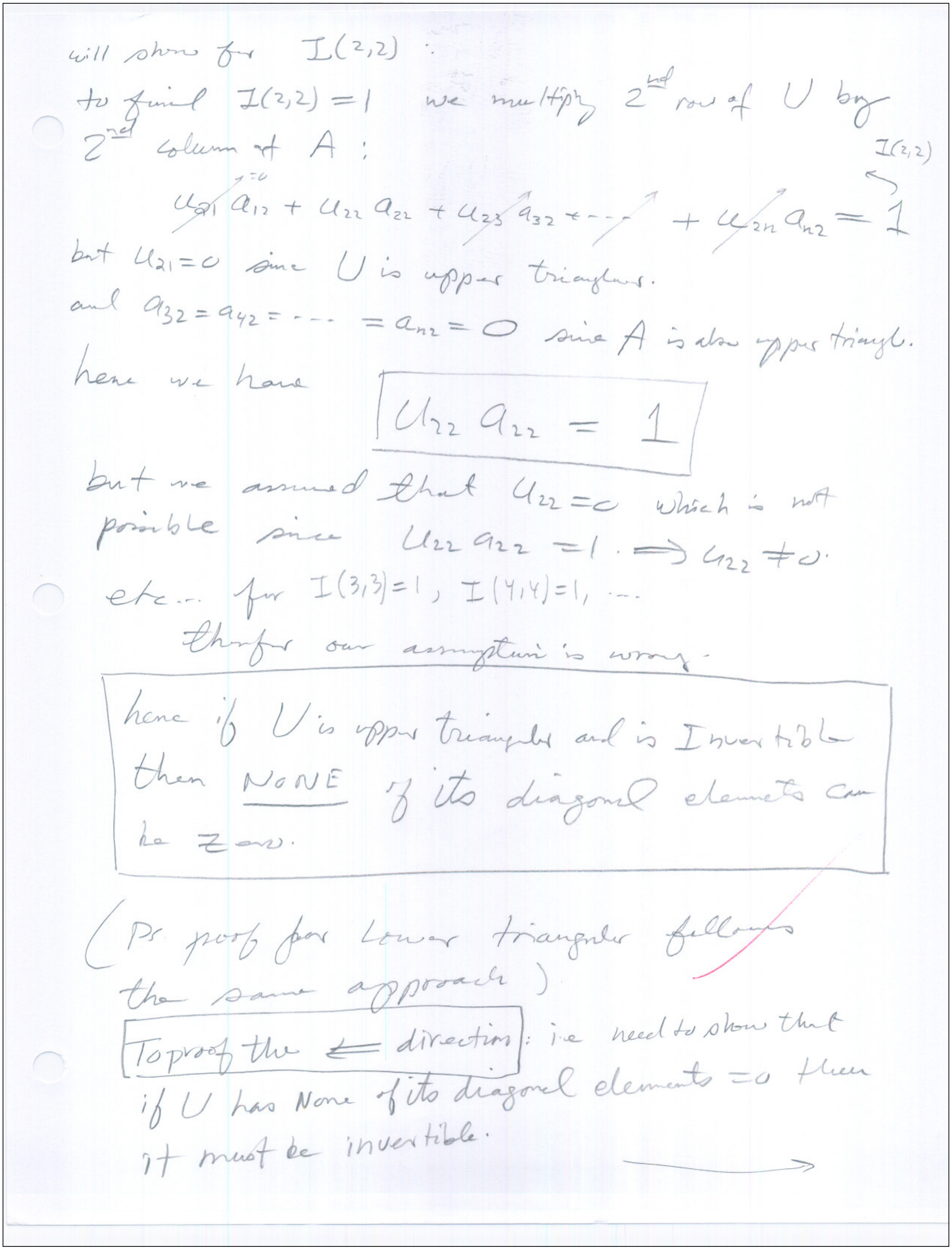
hence ① becomes $u_{11}a_{11}=1$.

but we assumed that $u_{11}=0$ ⟹ Not possible since $u_{11}a_{11}=1$.

hence $\boxed{u_{11} \text{ can Not be zero.}}$

Now do the same thing to get an equation for $I(2,2)$. this equation will show that u_{22} Can Not be zero.

we continue this way to show that u_{ii} can NOT be zero ⟶



will show for $I(2,2)$:

to find $I(2,2)=1$ we multiply 2^{nd} row of U by 2^{nd} column of A :

$I(2,2)$

$$\overset{=0}{\cancel{u_{21}}}\, a_{12} + u_{22}a_{22} + u_{23}\cancel{a_{32}} + \cancel{\cdots} + u_{2n}\cancel{a_{n2}} = 1$$

but $u_{21}=0$ since U is upper triangular.

and $a_{32}=a_{42}=\cdots=a_{n2}=0$ since A is also upper triangl.

hence we have $\boxed{u_{22}a_{22} = 1}$

but we assumed that $u_{22}=0$ which is not possible since $u_{22}a_{22}=1 \Rightarrow u_{22}\neq 0$.

etc... for $I(3,3)=1$, $I(4,4)=1$, ...

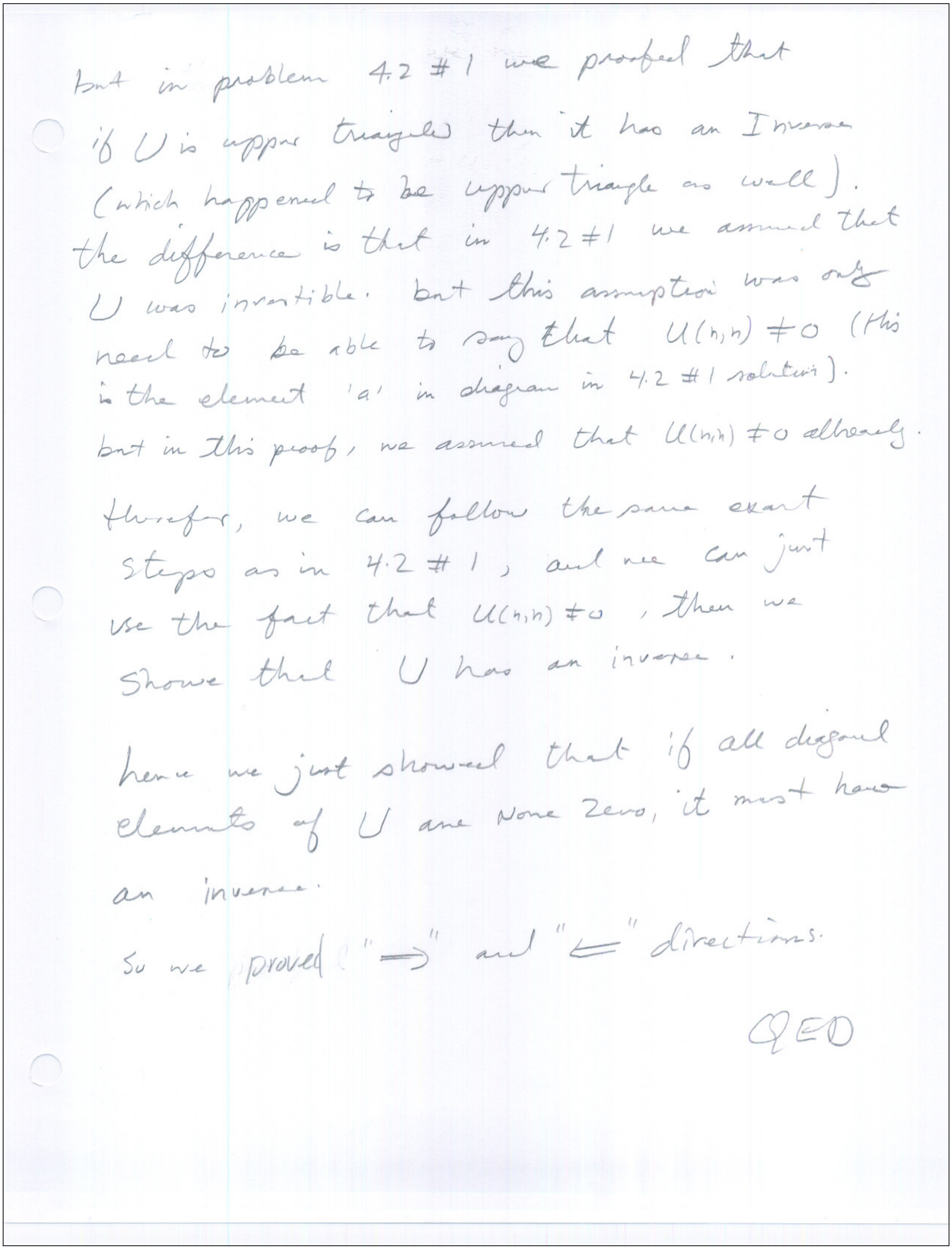
therefor our assumption is wrong.

hence if U is upper triangular and is Invertible then NONE of its diagonal elements can be zero.

(Ps. proof for Lower triangular follows the same approach)

To proof the $\Leftarrow$ direction: i.e need to show that if U has None of its diagonal elements $=0$ then it must be invertible.

→



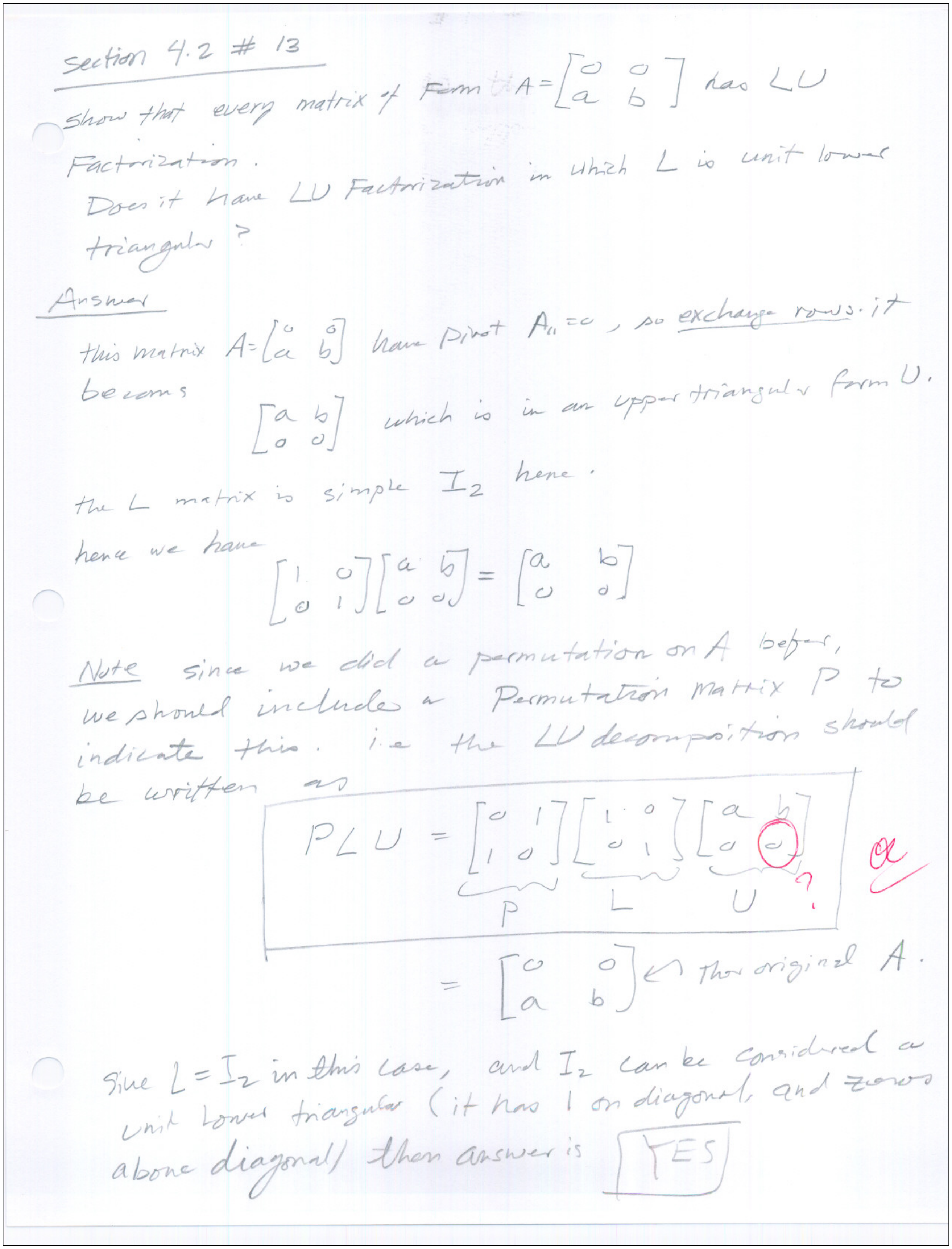
but in problem 4.2 #1 we proofed that if U is upper triangle then it has an Inverse (which happened to be upper triangle as well). the difference is that in 4.2 #1 we assumed that U was invertible. but this assumption was only need to be able to say that $U(n,n) \neq 0$ (this is the element 'a' in diagram in 4.2 #1 solution). but in this proof, we assumed that $U(n,n) \neq 0$ already.

therefore, we can follow the same exact steps as in 4.2 #1, and we can just use the fact that $U(n,n) \neq 0$, then we showe that U has an inverse.

hence we just showed that if all diagonal elements of U are None zero, it must have an inverse.

So we proved "$\Longrightarrow$" and "$\Longleftarrow$" directions.

QED



section 4.2 # 13

Show that every matrix of Form $A = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix}$ has LU Factorization.

Does it have LU Factorization in which L is unit lower triangular?

Answer

this matrix $A = \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix}$ have Pivot $A_{11} = 0$, so exchange rows. it becoms

$\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$ which is in an upper triangular form U.

the L matrix is simple I_2 here.

hence we have

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}$$

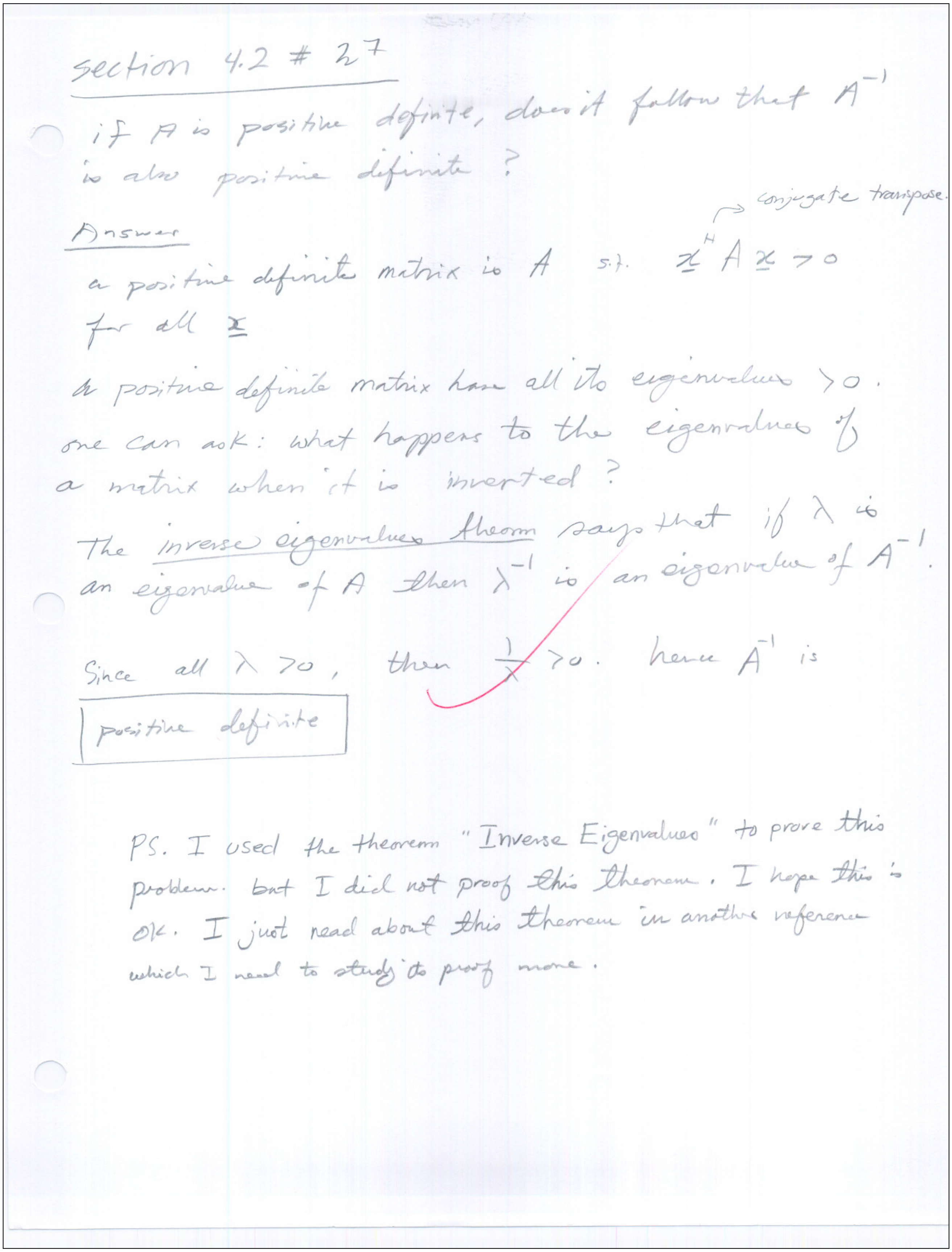
Note since we did a permutation on A befor, we should include a Permutation matrix P to indicate this. i.e the LU decomposition should be written as

$$\boxed{PLU = \underbrace{\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}}_{L} \underbrace{\begin{bmatrix} a & b \\ 0 & 0 \end{bmatrix}}_{U} \; ?}$$

a

$$= \begin{bmatrix} 0 & 0 \\ a & b \end{bmatrix} \leftarrow \text{the original } A.$$

Since $L = I_2$ in this case, and I_2 can be considered a unit Lower triangular (it has 1 on diagonal, and zeros above diagonal) then answer is $\boxed{\text{YES}}$



section 4.2 # 27

if A is positive definite, does it follow that A^{-1} is also positive definite ?

Answer

conjugate transpose.

a positive definite matrix is A s.t. $\underline{x}^H A \underline{x} > 0$ for all $\underline{x}$

a positive definite matrix have all its eigenvalues > 0.

one can ask: what happens to the eigenvalues of a matrix when it is inverted ?

The inverse eigenvalues theorm says that if λ is an eigenvalue of A then λ^{-1} is an eigenvalue of A^{-1}.

Since all $\lambda > 0$, then $\frac{1}{\lambda} > 0$. hence A^{-1} is

$\boxed{\text{positive definite}}$

PS. I used the theorem "Inverse Eigenvalues" to prove this problem. but I did not proof this theorem. I hope this is ok. I just read about this theorem in another reference which I need to study its proof more.

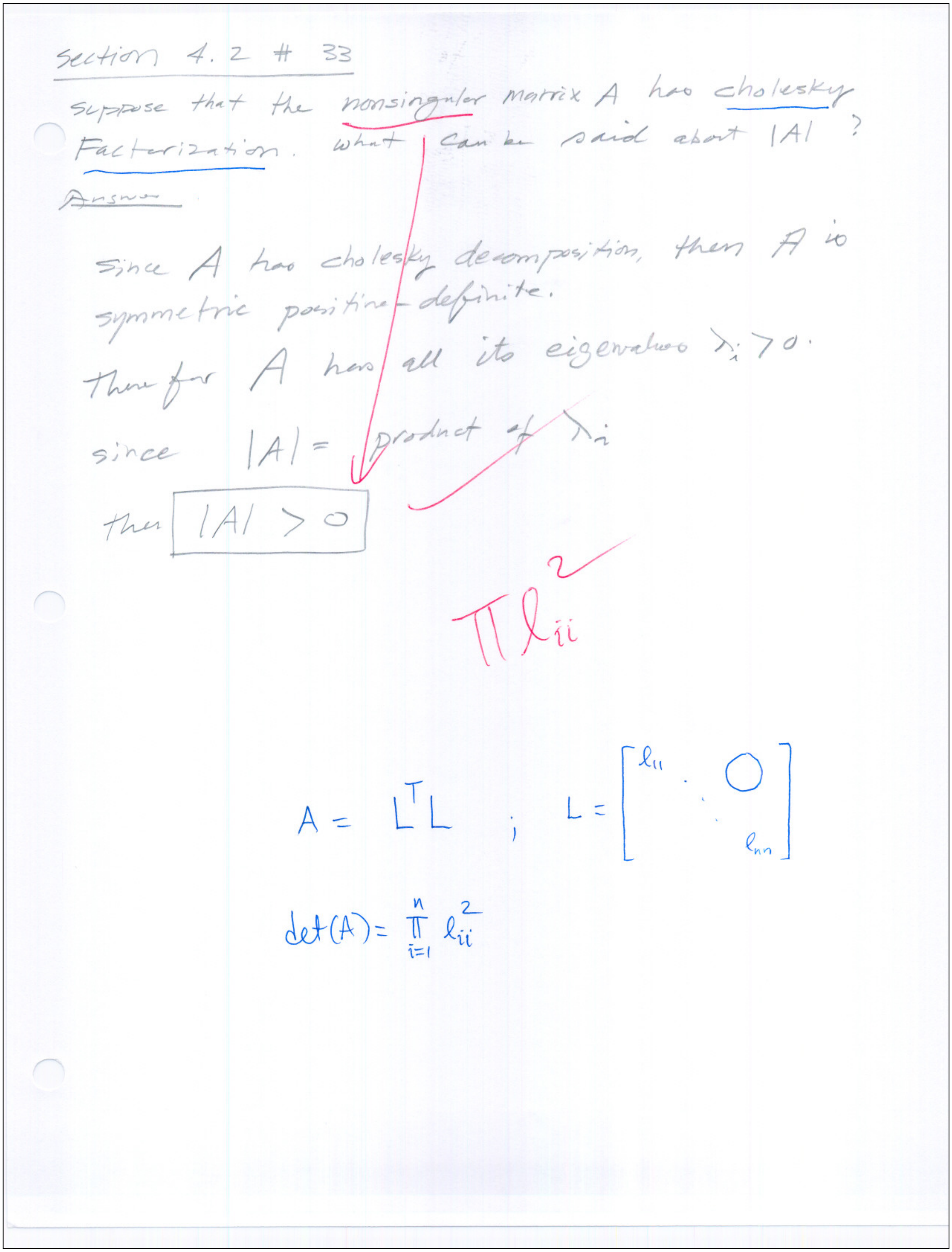
Section 4.2 #30

Find LU Factorization of $A = \begin{bmatrix} 3 & 0 & 1 \\ 0 & -1 & 3 \\ 1 & 3 & 0 \end{bmatrix}$

$$\begin{bmatrix} \overset{\text{pivot}}{\boxed{3}} & 0 & 1 \\ 0 & -1 & 3 \\ 1 & 3 & 0 \end{bmatrix} \xrightarrow{l_{31} = \frac{1}{3}} \begin{bmatrix} 3 & 0 & 1 \\ 0 & \overset{\text{pivot}}{\boxed{-1}} & 3 \\ 0 & 3 & -\frac{1}{3} \end{bmatrix} \xrightarrow{l_{32} = -3} \begin{bmatrix} 3 & 0 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & \frac{26}{3} \end{bmatrix} \leftarrow U$$

$$\text{so } L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ l_{31} & l_{32} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \frac{1}{3} & -3 & 1 \end{bmatrix} \leftarrow L$$

$$\text{So} \quad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -\frac{1}{3} & -3 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 & 1 \\ 0 & -1 & 3 \\ 0 & 0 & \frac{26}{3} \end{bmatrix}$$



section 4.2 # 33

suppose that the nonsingular matrix A has cholesky Factorization. what can be said about |A| ?

Answer

Since A has cholesky decomposition, then A is symmetric positive definite.

Therefor A has all its eigenvalues $\lambda_i > 0$.

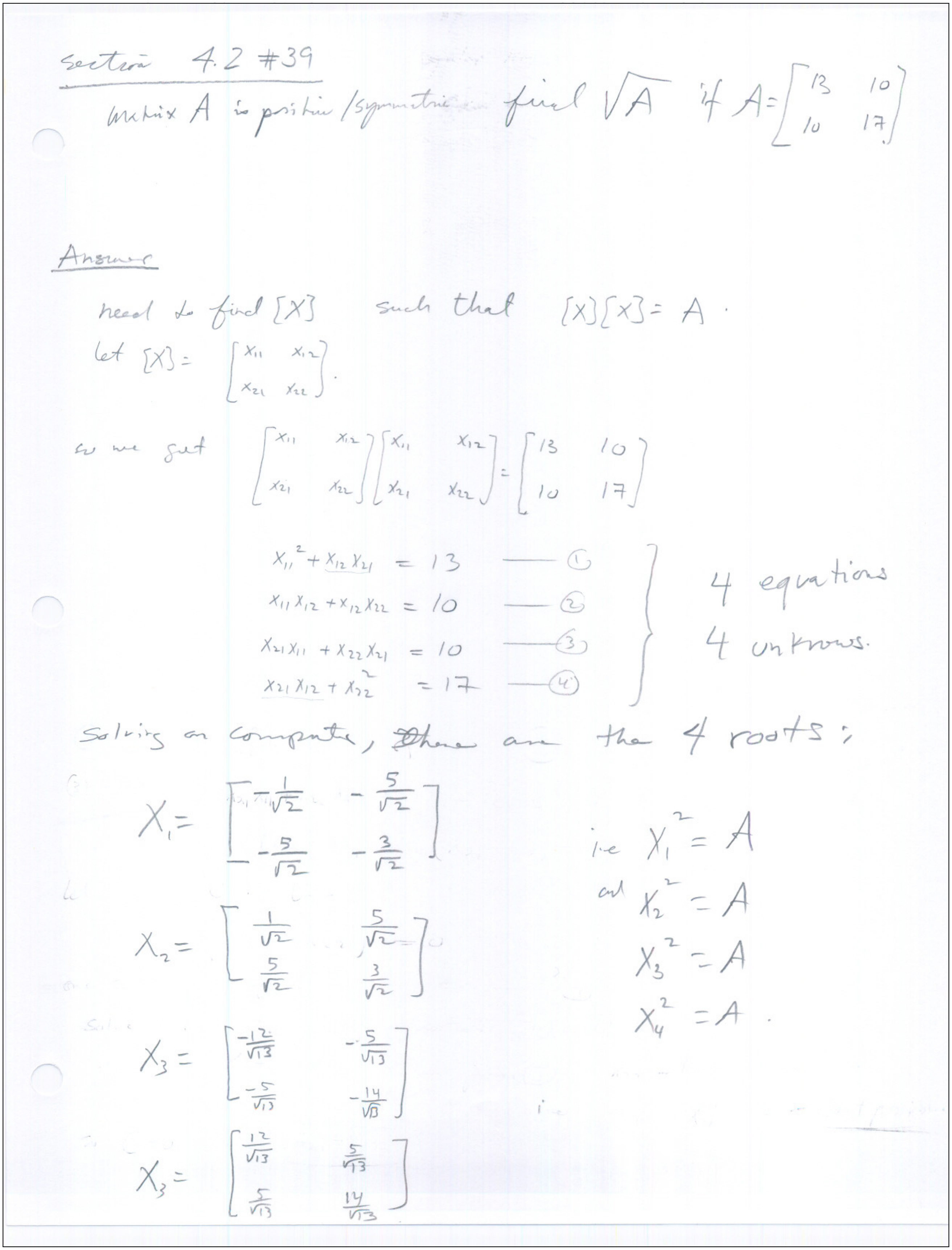
since $|A| =$ product of λ_i

then $\boxed{|A| > 0}$

$\prod l_{ii}^2$

$A = L^T L$; $L = \begin{bmatrix} l_{11} & & 0 \\ & \ddots & \\ & & l_{nn} \end{bmatrix}$

$\det(A) = \prod_{i=1}^{n} l_{ii}^2$



section 4.2 #39

matrix A is positive/symmetric find $\sqrt{A}$ if $A=\begin{bmatrix}13 & 10\\ 10 & 17\end{bmatrix}$

Answer

need to find $[X]$ such that $[X][X]=A$.

let $[X]=\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}$.

so we get $\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}\begin{bmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{bmatrix}=\begin{bmatrix}13 & 10\\ 10 & 17\end{bmatrix}$

$$\left.\begin{array}{ll} x_{11}^2+x_{12}x_{21}=13 & \text{——} (1)\\ x_{11}x_{12}+x_{12}x_{22}=10 & \text{——} (2)\\ x_{21}x_{11}+x_{22}x_{21}=10 & \text{——} (3)\\ x_{21}x_{12}+x_{22}^2=17 & \text{——} (4)\end{array}\right\}$$

4 equations
4 unknows.

Solving on computer, ~~there~~ are the 4 roots:

$$X_1=\begin{bmatrix}-\frac{1}{\sqrt{2}} & -\frac{5}{\sqrt{2}}\\ -\frac{5}{\sqrt{2}} & -\frac{3}{\sqrt{2}}\end{bmatrix}$$

$$X_2=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{5}{\sqrt{2}}\\ \frac{5}{\sqrt{2}} & \frac{3}{\sqrt{2}}\end{bmatrix}$$

$$X_3=\begin{bmatrix}\frac{-12}{\sqrt{13}} & -\frac{5}{\sqrt{13}}\\ \frac{-5}{\sqrt{13}} & -\frac{14}{\sqrt{13}}\end{bmatrix}$$

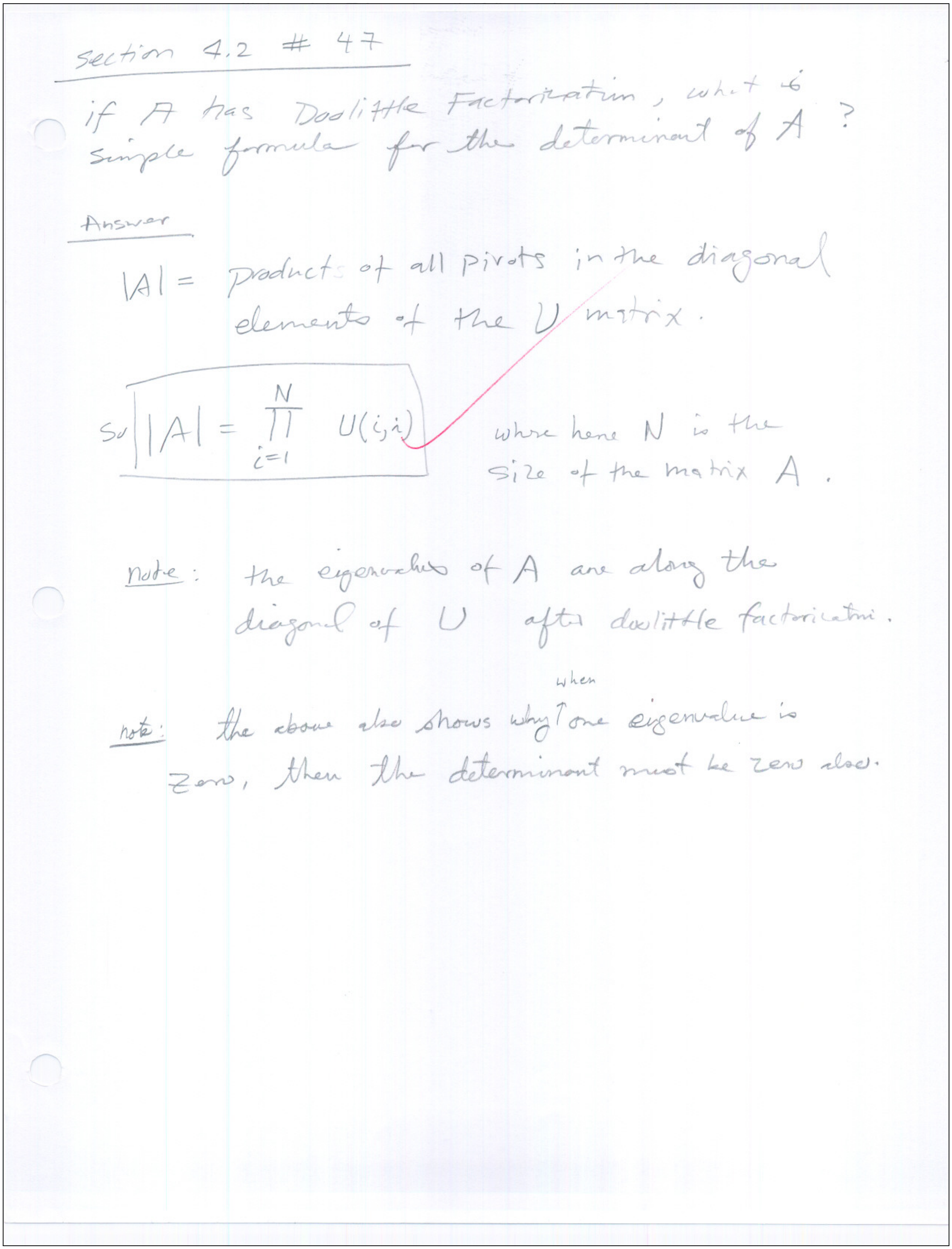
$$X_3=\begin{bmatrix}\frac{12}{\sqrt{13}} & \frac{5}{\sqrt{13}}\\ \frac{5}{\sqrt{13}} & \frac{14}{\sqrt{13}}\end{bmatrix}$$

i.e $X_1^2=A$

and $X_2^2=A$

$X_3^2=A$

$X_4^2=A$.



section 4.2 # 47

if A has Doolittle Factorizatin, what is simple formula for the determinant of A ?

Answer

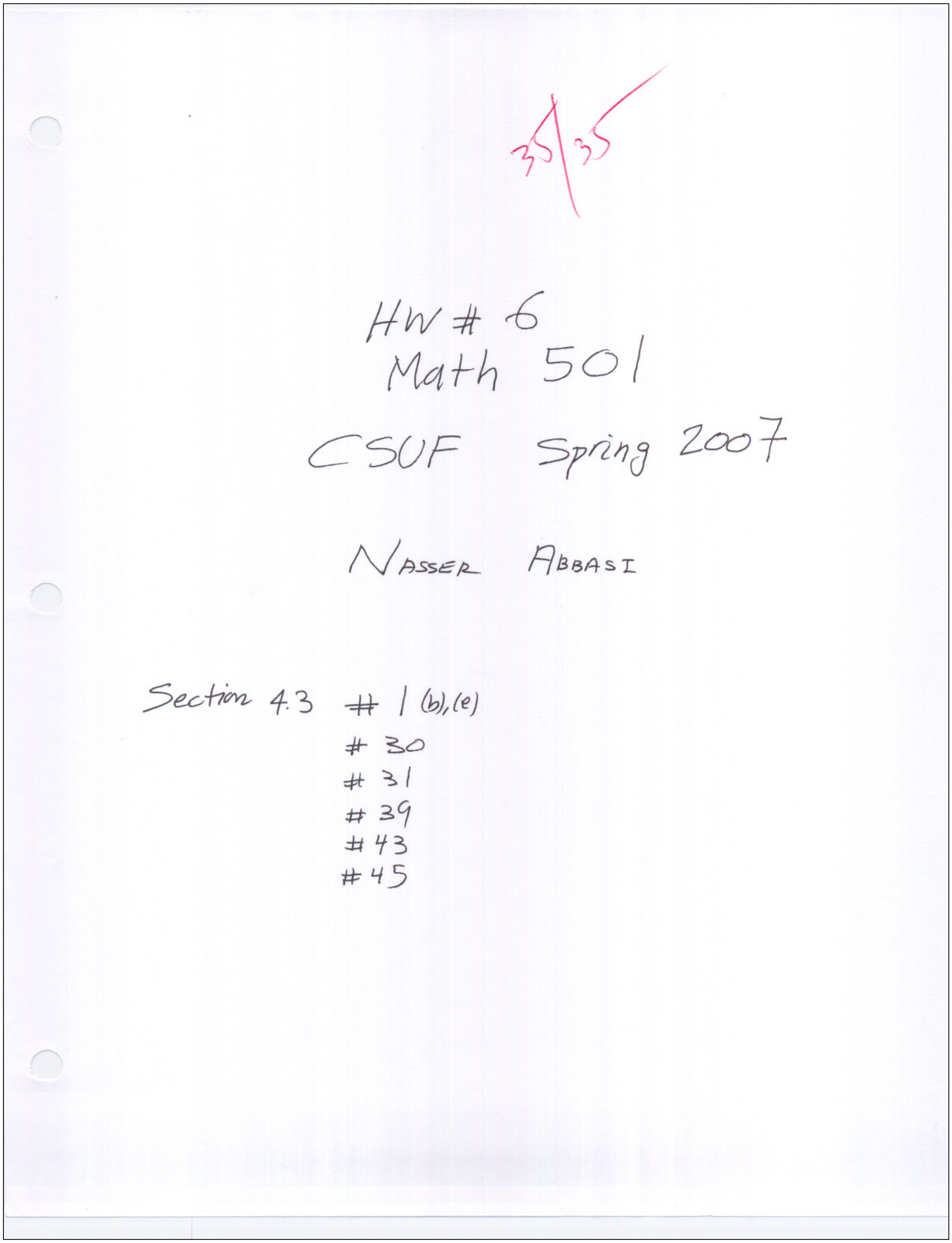
$|A| =$ products of all pivots in the diagonal elements of the U matrix.

So $\boxed{|A| = \prod_{i=1}^{N} U(i,i)}$ where here N is the size of the matrix A.

note: the eigenvalues of A are along the diagonal of U after doolittle factorizatin.

note: the above also shows why when one eigenvalue is zero, then the determinant must be zero also.

4.7 HW 6



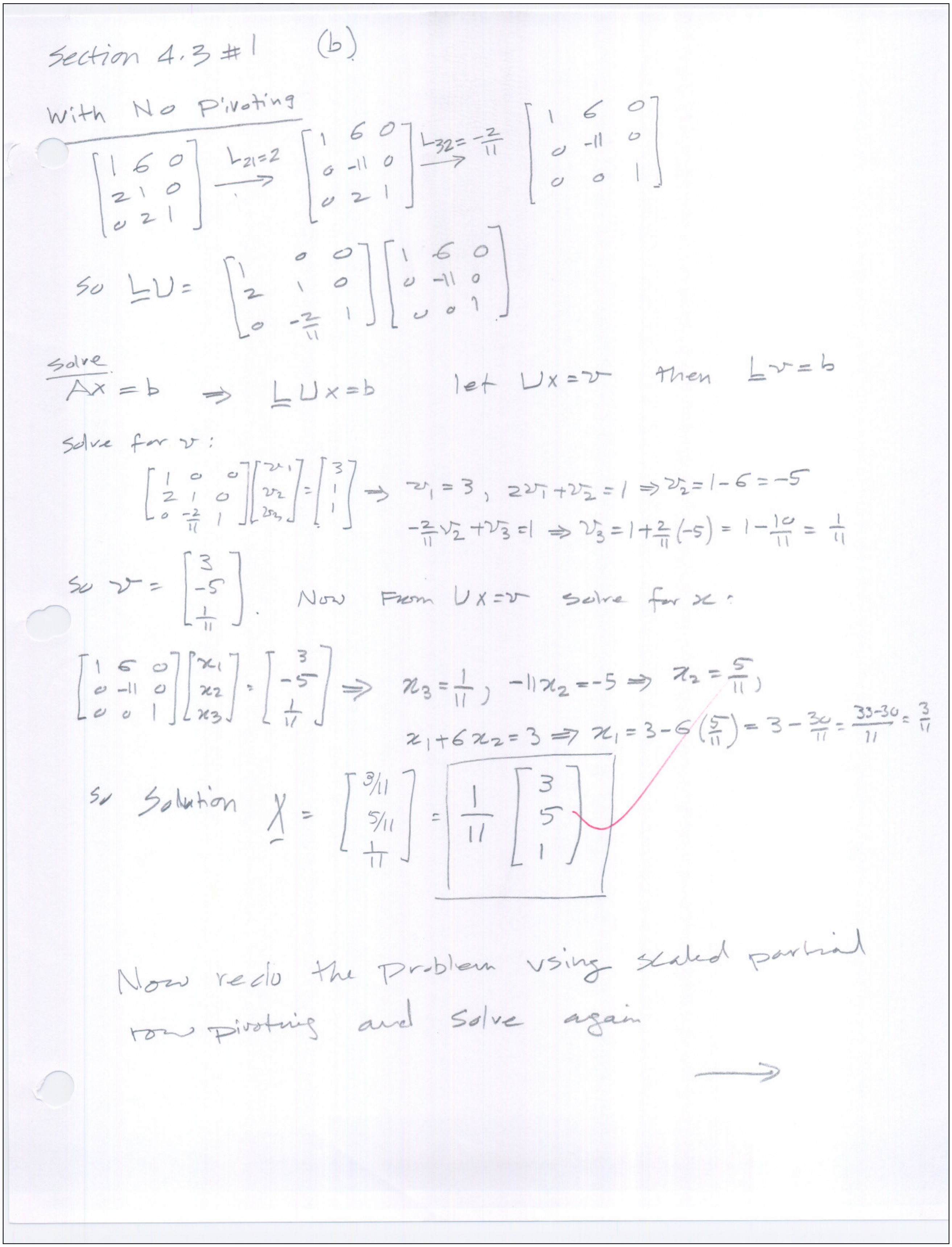
35/35

HW # 6
Math 501

CSUF Spring 2007

NASSER ABBASI

Section 4.3 # 1 (b),(e)
30
31
39
43
45



Section 4.3 #1 (b)

With No Pivoting

$$\begin{bmatrix}1&6&0\\2&1&0\\0&2&1\end{bmatrix} \xrightarrow{L_{21}=2} \begin{bmatrix}1&6&0\\0&-11&0\\0&2&1\end{bmatrix} \xrightarrow{L_{32}=-\frac{2}{11}} \begin{bmatrix}1&6&0\\0&-11&0\\0&0&1\end{bmatrix}$$

$$\text{so } \underline{L}\underline{U} = \begin{bmatrix}1&0&0\\2&1&0\\0&-\frac{2}{11}&1\end{bmatrix}\begin{bmatrix}1&6&0\\0&-11&0\\0&0&1\end{bmatrix}$$

solve
$Ax=b \Rightarrow \underline{L}\underline{U}x=b$ let $\underline{U}x=v$ then $\underline{L}v=b$

solve for v:

$$\begin{bmatrix}1&0&0\\2&1&0\\0&-\frac{2}{11}&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}=\begin{bmatrix}3\\1\\1\end{bmatrix} \Rightarrow v_1=3,\ 2v_1+v_2=1 \Rightarrow v_2=1-6=-5$$

$$-\frac{2}{11}v_2+v_3=1 \Rightarrow v_3=1+\frac{2}{11}(-5)=1-\frac{10}{11}=\frac{1}{11}$$

$$\text{so } v=\begin{bmatrix}3\\-5\\\frac{1}{11}\end{bmatrix}.$$ Now From $\underline{U}x=v$ solve for x:

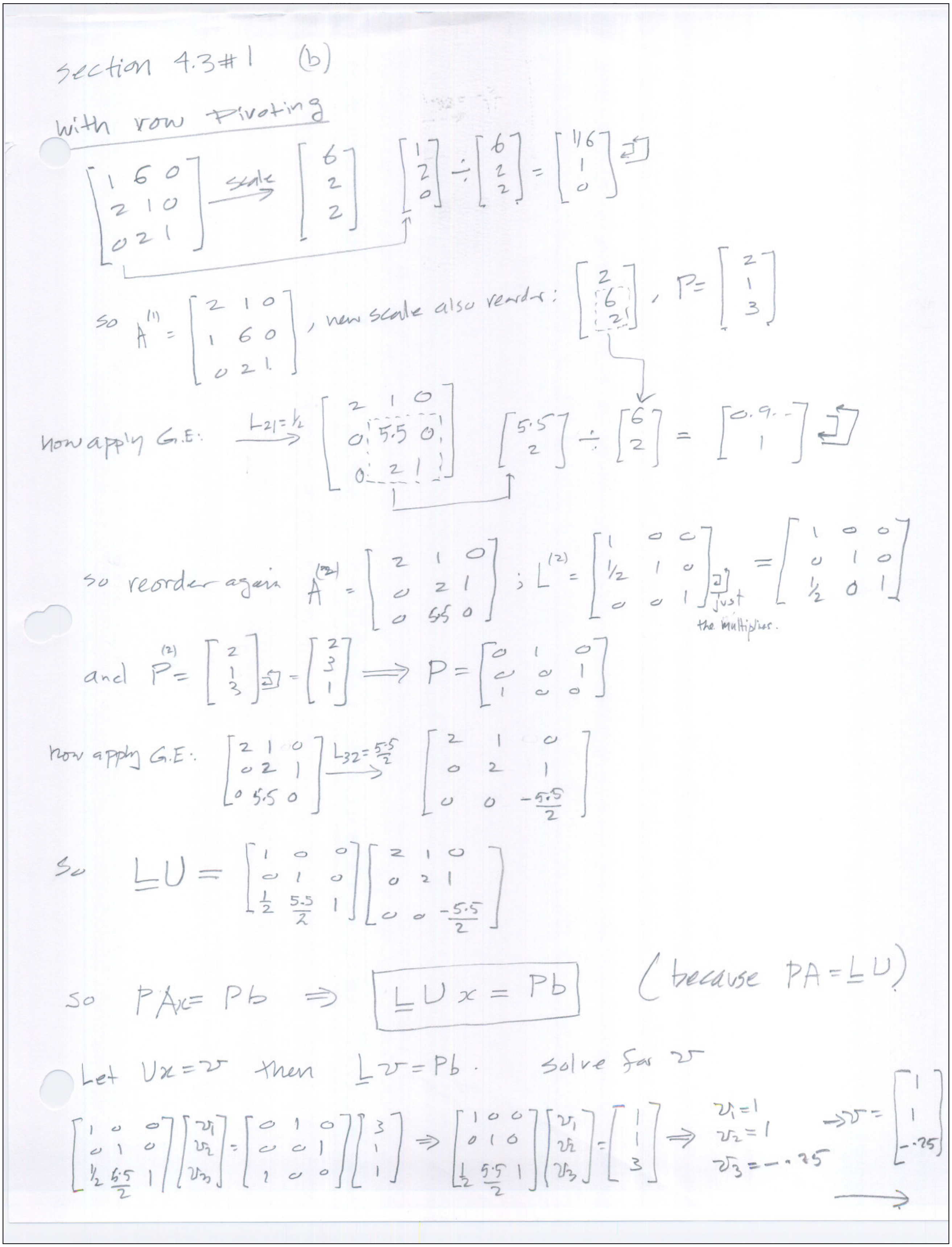
$$\begin{bmatrix}1&6&0\\0&-11&0\\0&0&1\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}3\\-5\\\frac{1}{11}\end{bmatrix} \Rightarrow x_3=\frac{1}{11},\ -11x_2=-5 \Rightarrow x_2=\frac{5}{11},$$

$$x_1+6x_2=3 \Rightarrow x_1=3-6\left(\frac{5}{11}\right)=3-\frac{30}{11}=\frac{33-30}{11}=\frac{3}{11}$$

$$\text{so Solution } \underline{X}=\begin{bmatrix}3/11\\5/11\\\frac{1}{11}\end{bmatrix}=\boxed{\frac{1}{11}\begin{bmatrix}3\\5\\1\end{bmatrix}}$$

Now redo the problem using scaled partial row pivoting and solve again

→



section 4.3#1 (b)

with row Pivoting

$$\begin{bmatrix}1&6&0\\2&1&0\\0&2&1\end{bmatrix}\xrightarrow{\text{scale}}\begin{bmatrix}6\\2\\2\end{bmatrix}\quad\begin{bmatrix}1\\2\\0\end{bmatrix}\div\begin{bmatrix}6\\2\\2\end{bmatrix}=\begin{bmatrix}1/6\\1\\0\end{bmatrix}\rightleftarrows$$

so $A^{(1)}=\begin{bmatrix}2&1&0\\1&6&0\\0&2&1\end{bmatrix}$, new scale also reorder: $\begin{bmatrix}2\\6\\2\end{bmatrix}$, $P=\begin{bmatrix}2\\1\\3\end{bmatrix}$

now apply G.E.: $\xrightarrow{L_{21}=1/2}\begin{bmatrix}2&1&0\\0&5.5&0\\0&2&1\end{bmatrix}\quad\begin{bmatrix}5.5\\2\end{bmatrix}\div\begin{bmatrix}6\\2\end{bmatrix}=\begin{bmatrix}0.9..\\1\end{bmatrix}\rightleftarrows$

so reorder again $A^{(2)}=\begin{bmatrix}2&1&0\\0&2&1\\0&5.5&0\end{bmatrix}$; $L^{(2)}=\begin{bmatrix}1&0&0\\1/2&1&0\\0&0&1\end{bmatrix}\rightleftarrows=\begin{bmatrix}1&0&0\\0&1&0\\1/2&0&1\end{bmatrix}$ just the multiplier.

and $P^{(2)}=\begin{bmatrix}2\\1\\3\end{bmatrix}\rightleftarrows=\begin{bmatrix}2\\3\\1\end{bmatrix}\Longrightarrow P=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}$

now apply G.E.: $\begin{bmatrix}2&1&0\\0&2&1\\0&5.5&0\end{bmatrix}\xrightarrow{L_{32}=\frac{5.5}{2}}\begin{bmatrix}2&1&0\\0&2&1\\0&0&-\frac{5.5}{2}\end{bmatrix}$

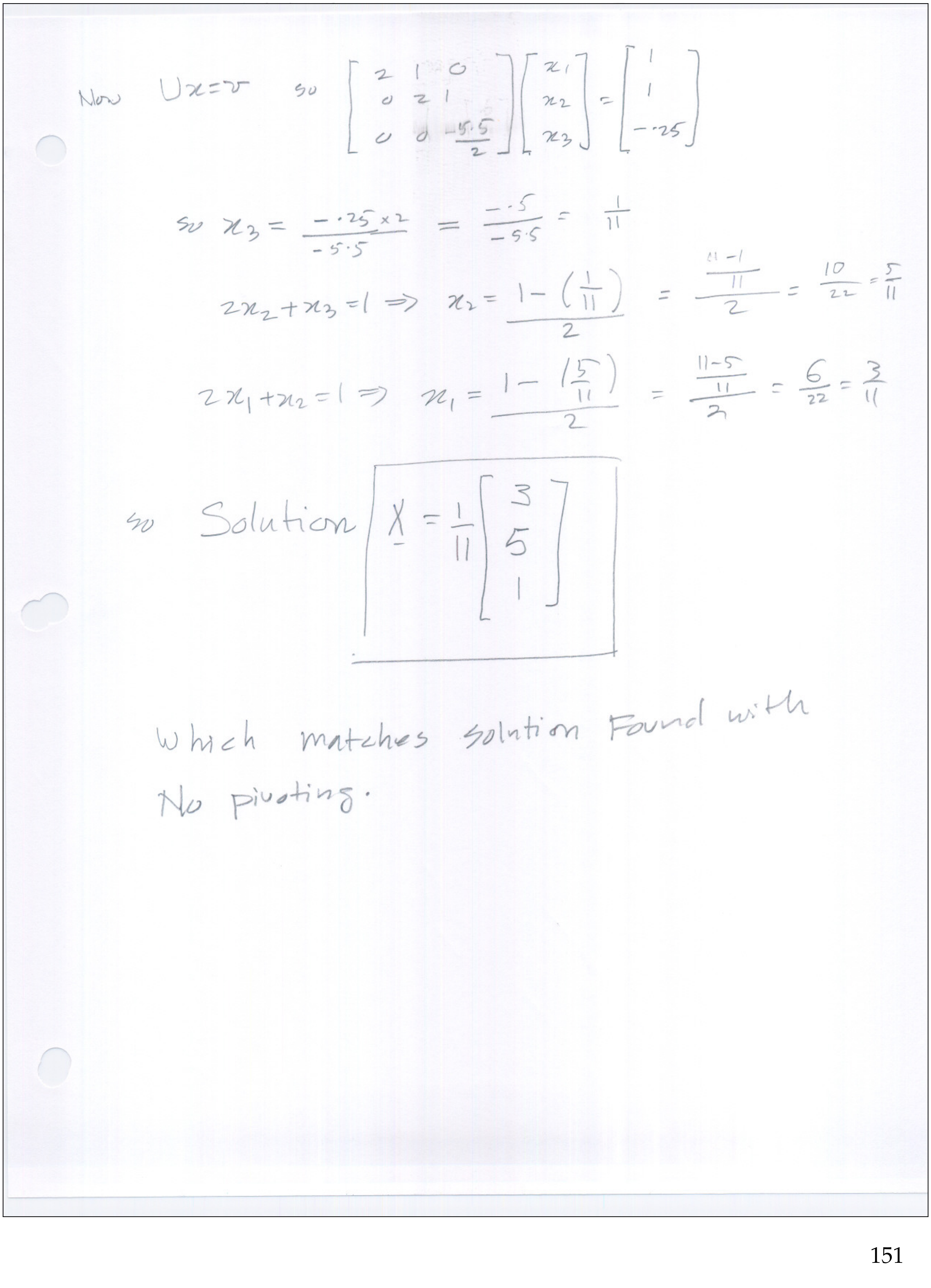
So $\underline{L}U=\begin{bmatrix}1&0&0\\0&1&0\\\frac{1}{2}&\frac{5.5}{2}&1\end{bmatrix}\begin{bmatrix}2&1&0\\0&2&1\\0&0&-\frac{5.5}{2}\end{bmatrix}$

so $PAx=Pb\Rightarrow\boxed{\underline{L}Ux=Pb}$ (because $PA=\underline{L}U$)

Let $Ux=v$ then $\underline{L}v=Pb$. solve for v

$$\begin{bmatrix}1&0&0\\0&1&0\\1/2&\frac{5.5}{2}&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}=\begin{bmatrix}0&1&0\\0&0&1\\1&0&0\end{bmatrix}\begin{bmatrix}3\\1\\1\end{bmatrix}\Rightarrow\begin{bmatrix}1&0&0\\0&1&0\\\frac{1}{2}&\frac{5.5}{2}&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\end{bmatrix}=\begin{bmatrix}1\\1\\3\end{bmatrix}\Rightarrow\begin{matrix}v_1=1\\v_2=1\\v_3=-.75\end{matrix}\rightarrow v=\begin{bmatrix}1\\1\\-.75\end{bmatrix}$$

$\longrightarrow$



Now $Ux=v$ so $\begin{bmatrix}2&1&0\\0&2&1\\0&0&\frac{-5.5}{2}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\end{bmatrix}=\begin{bmatrix}1\\1\\-.25\end{bmatrix}$

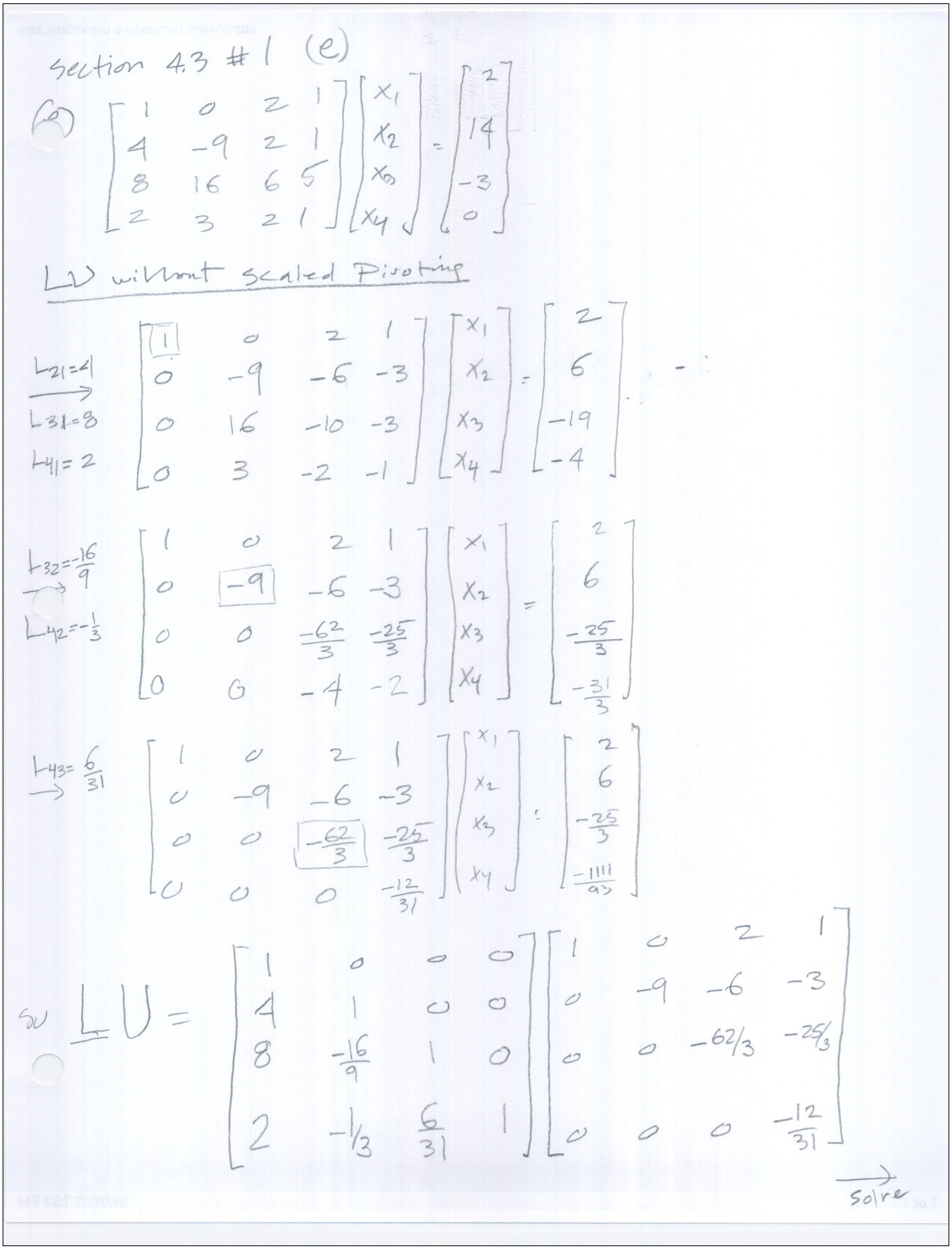
so $x_3 = \frac{-.25\times 2}{-5.5} = \frac{-.5}{-5.5} = \frac{1}{11}$

$2x_2+x_3=1 \Rightarrow x_2 = \frac{1-\left(\frac{1}{11}\right)}{2} = \frac{\frac{11-1}{11}}{2} = \frac{10}{22} = \frac{5}{11}$

$2x_1+x_2=1 \Rightarrow x_1 = \frac{1-\left(\frac{5}{11}\right)}{2} = \frac{\frac{11-5}{11}}{2} = \frac{6}{22} = \frac{3}{11}$

so Solution $\boxed{\underline{X} = \frac{1}{11}\begin{bmatrix}3\\5\\1\end{bmatrix}}$

Which matches solution Found with No pivoting.



section 4.3 #1 (e)

(e)
$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 4 & -9 & 2 & 1 \\ 8 & 16 & 6 & 5 \\ 2 & 3 & 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 14 \\ -3 \\ 0 \end{bmatrix}$$

LU without scaled Pivoting

$L_{21}=4$ $\longrightarrow$
$L_{31}=8$
$L_{41}=2$

$$\begin{bmatrix} \boxed{1} & 0 & 2 & 1 \\ 0 & -9 & -6 & -3 \\ 0 & 16 & -10 & -3 \\ 0 & 3 & -2 & -1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ -19 \\ -4 \end{bmatrix}$$

$L_{32}=-\frac{16}{9}$ $\longrightarrow$
$L_{42}=-\frac{1}{3}$

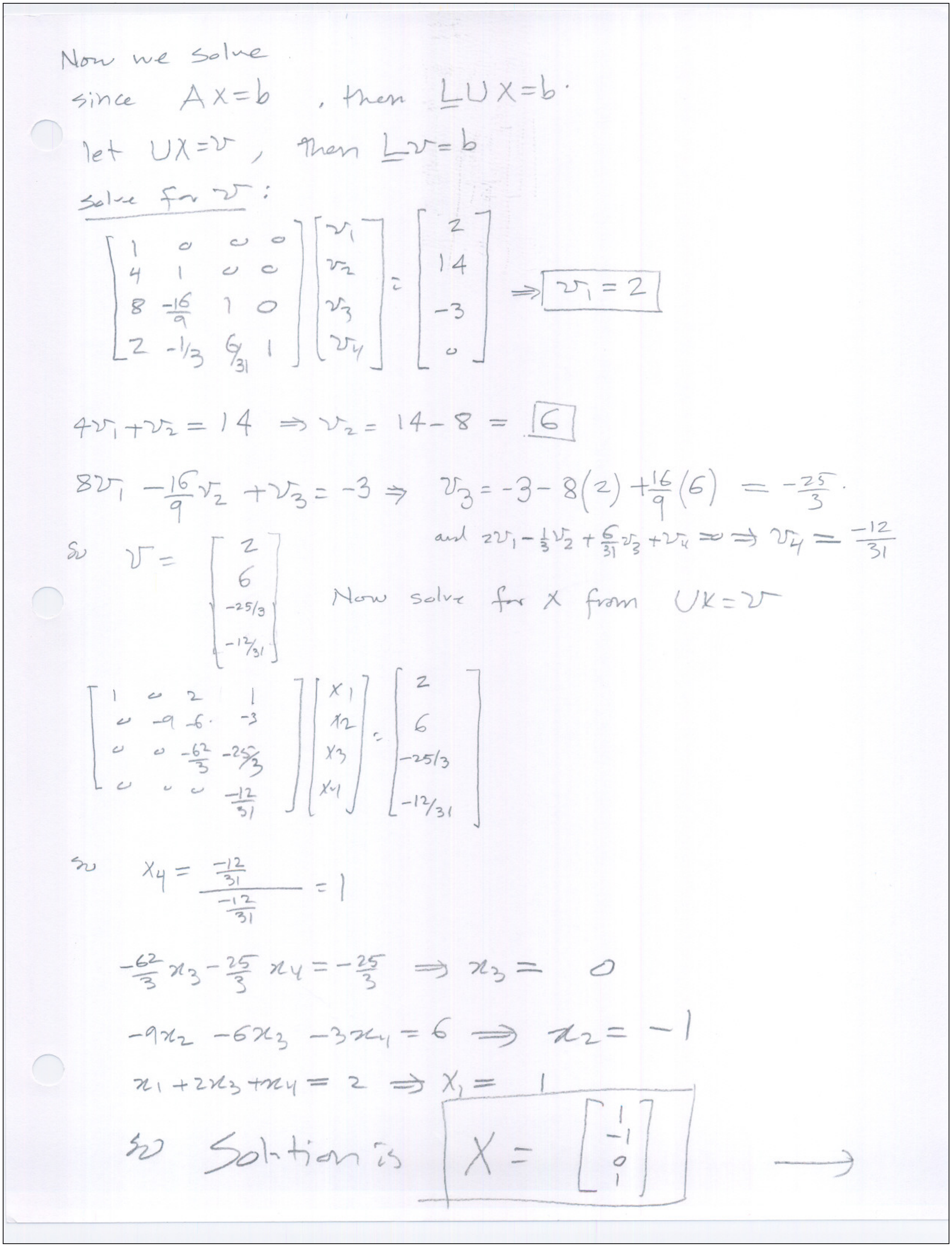
$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & \boxed{-9} & -6 & -3 \\ 0 & 0 & \frac{-62}{3} & \frac{-25}{3} \\ 0 & 0 & -4 & -2 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ -\frac{25}{3} \\ -\frac{31}{3} \end{bmatrix}$$

$L_{43}=\frac{6}{31}$ $\longrightarrow$

$$\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & -9 & -6 & -3 \\ 0 & 0 & \boxed{\frac{-62}{3}} & \frac{-25}{3} \\ 0 & 0 & 0 & \frac{-12}{31} \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 2 \\ 6 \\ -\frac{25}{3} \\ \frac{-1111}{93} \end{bmatrix}$$

so
$$LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 4 & 1 & 0 & 0 \\ 8 & -\frac{16}{9} & 1 & 0 \\ 2 & -1/3 & \frac{6}{31} & 1 \end{bmatrix}\begin{bmatrix} 1 & 0 & 2 & 1 \\ 0 & -9 & -6 & -3 \\ 0 & 0 & -62/3 & -25/3 \\ 0 & 0 & 0 & \frac{-12}{31} \end{bmatrix}$$

$\longrightarrow$ solve



Now we solve

since $Ax=b$, then $\underline{L}UX=b$.

let $UX=v$, then $\underline{L}v=b$

<u>solve for v:</u>

$$\begin{bmatrix}1&0&0&0\\4&1&0&0\\8&-\frac{16}{9}&1&0\\2&-1/3&6/31&1\end{bmatrix}\begin{bmatrix}v_1\\v_2\\v_3\\v_4\end{bmatrix}=\begin{bmatrix}2\\14\\-3\\0\end{bmatrix}\Rightarrow\boxed{v_1=2}$$

$4v_1+v_2=14 \Rightarrow v_2=14-8=\boxed{6}$

$8v_1-\frac{16}{9}v_2+v_3=-3 \Rightarrow v_3=-3-8(2)+\frac{16}{9}(6)=-\frac{25}{3}.$

and $2v_1-\frac{1}{3}v_2+\frac{6}{31}v_3+v_4=0 \Rightarrow v_4=\frac{-12}{31}$

so $v=\begin{bmatrix}2\\6\\-25/3\\-12/31\end{bmatrix}$ Now solve for X from $UX=v$

$$\begin{bmatrix}1&0&2&1\\0&-9&-6&-3\\0&0&-\frac{62}{3}&-\frac{25}{3}\\0&0&0&\frac{-12}{31}\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}2\\6\\-25/3\\-12/31\end{bmatrix}$$

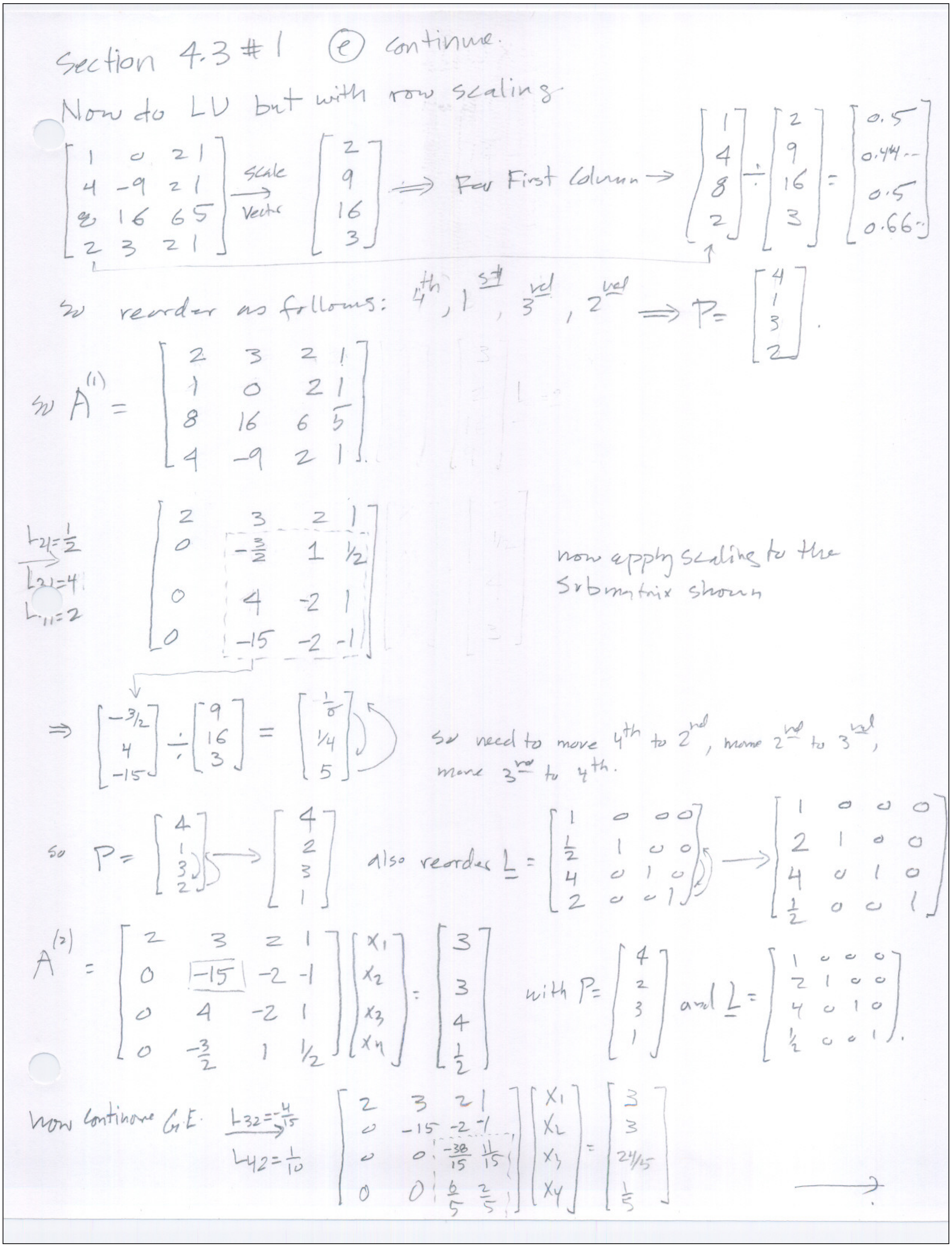
so $x_4=\dfrac{\frac{-12}{31}}{\frac{-12}{31}}=1$

$-\frac{62}{3}x_3-\frac{25}{3}x_4=-\frac{25}{3} \Rightarrow x_3=0$

$-9x_2-6x_3-3x_4=6 \Rightarrow x_2=-1$

$x_1+2x_3+x_4=2 \Rightarrow X_1=1$

so Solution is $\boxed{X=\begin{bmatrix}1\\-1\\0\\1\end{bmatrix}}$ ⟶



Section 4.3 #1 (e) continue.

Now do LU but with row scaling

$$\begin{bmatrix}1&0&2&1\\4&-9&2&1\\8&16&6&5\\2&3&2&1\end{bmatrix}\xrightarrow[\text{Vector}]{\text{scale}}\begin{bmatrix}2\\9\\16\\3\end{bmatrix}\Longrightarrow \text{For First Column}\rightarrow\begin{bmatrix}1\\4\\8\\2\end{bmatrix}\div\begin{bmatrix}2\\9\\16\\3\end{bmatrix}=\begin{bmatrix}0.5\\0.44\ldots\\0.5\\0.66\ldots\end{bmatrix}$$

so reorder as follows: $4^{th}, 1^{st}, 3^{rd}, 2^{nd} \Longrightarrow P=\begin{bmatrix}4\\1\\3\\2\end{bmatrix}$.

$$\text{so } A^{(1)}=\begin{bmatrix}2&3&2&1\\1&0&2&1\\8&16&6&5\\4&-9&2&1\end{bmatrix}.$$

$$\xrightarrow[\substack{L_{31}=4\\L_{41}=2}]{L_{21}=\frac{1}{2}}\begin{bmatrix}2&3&2&1\\0&-\frac{3}{2}&1&\frac12\\0&4&-2&1\\0&-15&-2&-1\end{bmatrix}$$

now apply scaling to the Submatrix shown

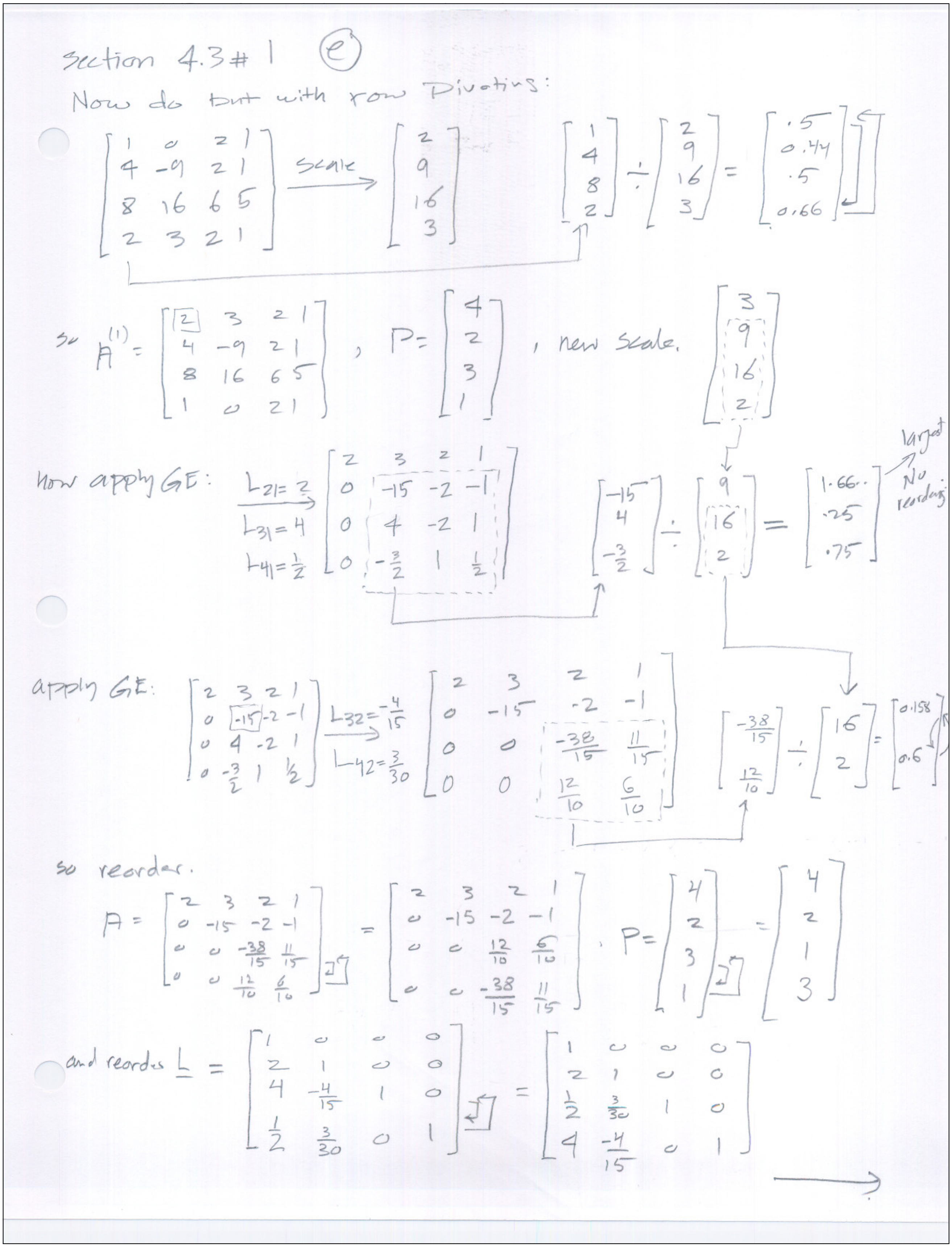
$$\Rightarrow\begin{bmatrix}-3/2\\4\\-15\end{bmatrix}\div\begin{bmatrix}9\\16\\3\end{bmatrix}=\begin{bmatrix}\frac{1}{6}\\1/4\\5\end{bmatrix}$$

so need to move 4^{th} to 2^{nd}, move 2^{nd} to 3^{rd}, move 3^{rd} to 4^{th}.

$$\text{so } P=\begin{bmatrix}4\\1\\3\\2\end{bmatrix}\longrightarrow\begin{bmatrix}4\\2\\3\\1\end{bmatrix}\quad \text{also reorder } \underline{L}=\begin{bmatrix}1&0&0&0\\\frac12&1&0&0\\4&0&1&0\\2&0&0&1\end{bmatrix}\longrightarrow\begin{bmatrix}1&0&0&0\\2&1&0&0\\4&0&1&0\\\frac12&0&0&1\end{bmatrix}$$

$$A^{(2)}=\begin{bmatrix}2&3&2&1\\0&\boxed{-15}&-2&-1\\0&4&-2&1\\0&-\frac32&1&\frac12\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}3\\3\\4\\\frac12\end{bmatrix}\quad\text{with } P=\begin{bmatrix}4\\2\\3\\1\end{bmatrix}\text{ and }\underline{L}=\begin{bmatrix}1&0&0&0\\2&1&0&0\\4&0&1&0\\\frac12&0&0&1\end{bmatrix}.$$

$$\text{now continue G.E. }\xrightarrow[L_{42}=\frac{1}{10}]{L_{32}=-\frac{4}{15}}\begin{bmatrix}2&3&2&1\\0&-15&-2&-1\\0&0&-\frac{38}{15}&\frac{11}{15}\\0&0&\frac65&\frac25\end{bmatrix}\begin{bmatrix}x_1\\x_2\\x_3\\x_4\end{bmatrix}=\begin{bmatrix}3\\3\\24/15\\\frac15\end{bmatrix}\longrightarrow$$



section 4.3 # 1 (e)

Now do but with row Pivoting:

$$\begin{bmatrix}1&0&2&1\\4&-9&2&1\\8&16&6&5\\2&3&2&1\end{bmatrix}\xrightarrow{\text{scale}}\begin{bmatrix}2\\9\\16\\3\end{bmatrix}\qquad\begin{bmatrix}1\\4\\8\\2\end{bmatrix}\div\begin{bmatrix}2\\9\\16\\3\end{bmatrix}=\begin{bmatrix}.5\\0.44\\.5\\0.66\end{bmatrix}$$

so $A^{(1)}=\begin{bmatrix}\boxed{2}&3&2&1\\4&-9&2&1\\8&16&6&5\\1&0&2&1\end{bmatrix}$, $P=\begin{bmatrix}4\\2\\3\\1\end{bmatrix}$, new scale. $\begin{bmatrix}3\\9\\16\\2\end{bmatrix}$

now apply GE:

$$\begin{matrix}L_{21}=2\\L_{31}=4\\L_{41}=\frac{1}{2}\end{matrix}\begin{bmatrix}2&3&2&1\\0&-15&-2&-1\\0&4&-2&1\\0&-\frac{3}{2}&1&\frac{1}{2}\end{bmatrix}\qquad\begin{bmatrix}-15\\4\\-\frac{3}{2}\end{bmatrix}\div\begin{bmatrix}9\\16\\2\end{bmatrix}=\begin{bmatrix}1.66..\\.25\\.75\end{bmatrix}$$

largest

No reordering

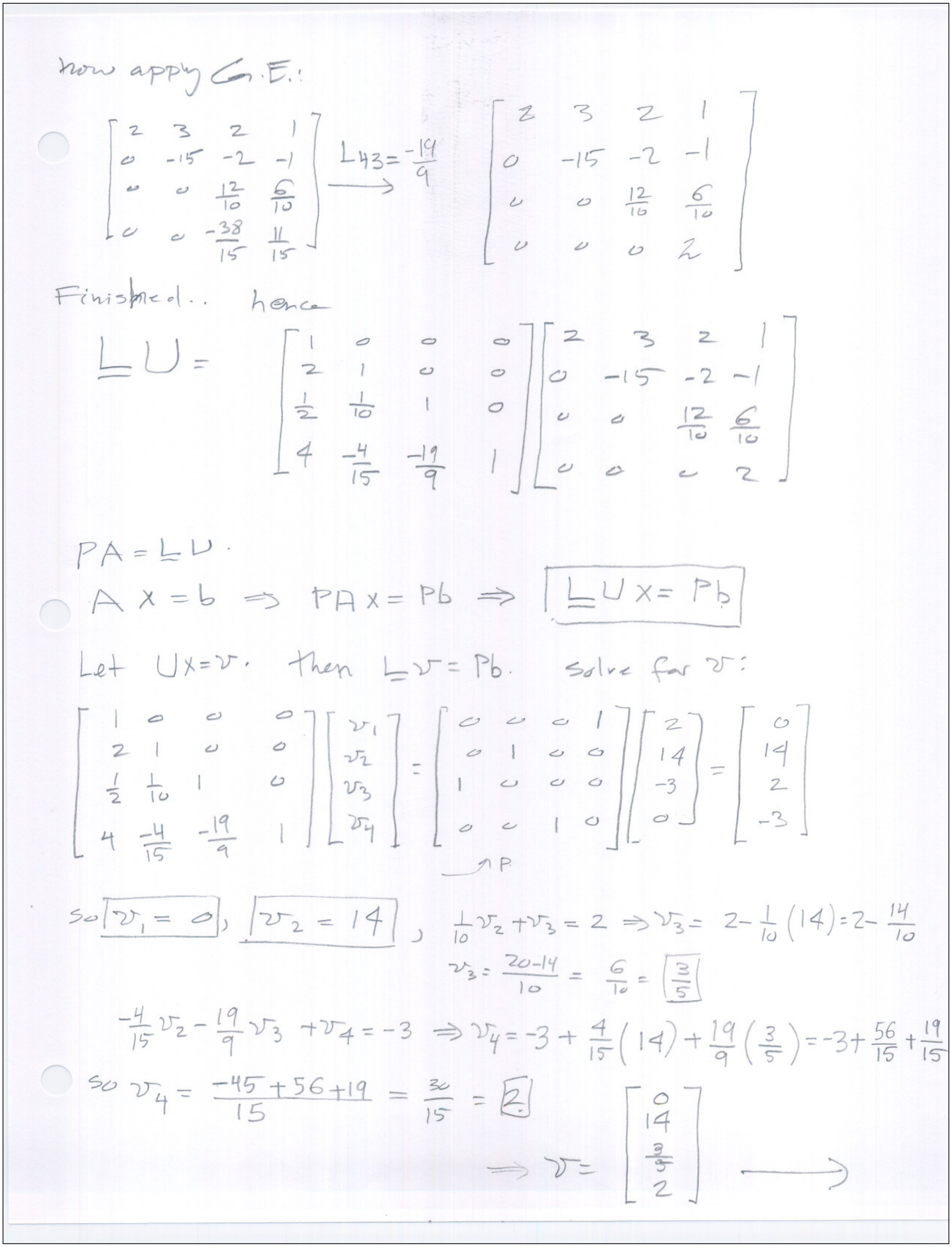
apply GE:

$$\begin{bmatrix}2&3&2&1\\0&\boxed{-15}&-2&-1\\0&4&-2&1\\0&-\frac{3}{2}&1&\frac{1}{2}\end{bmatrix}\xrightarrow[L_{42}=\frac{3}{30}]{L_{32}=-\frac{4}{15}}\begin{bmatrix}2&3&2&1\\0&-15&-2&-1\\0&0&-\frac{38}{15}&\frac{11}{15}\\0&0&\frac{12}{10}&\frac{6}{10}\end{bmatrix}\qquad\begin{bmatrix}-\frac{38}{15}\\\frac{12}{10}\end{bmatrix}\div\begin{bmatrix}16\\2\end{bmatrix}=\begin{bmatrix}0.158\\0.6\end{bmatrix}$$

so reorder.

$$A=\begin{bmatrix}2&3&2&1\\0&-15&-2&-1\\0&0&-\frac{38}{15}&\frac{11}{15}\\0&0&\frac{12}{10}&\frac{6}{10}\end{bmatrix}=\begin{bmatrix}2&3&2&1\\0&-15&-2&-1\\0&0&\frac{12}{10}&\frac{6}{10}\\0&0&-\frac{38}{15}&\frac{11}{15}\end{bmatrix},\quad P=\begin{bmatrix}4\\2\\3\\1\end{bmatrix}=\begin{bmatrix}4\\2\\1\\3\end{bmatrix}$$

and reorder $\underline{L}=\begin{bmatrix}1&0&0&0\\2&1&0&0\\4&-\frac{4}{15}&1&0\\\frac{1}{2}&\frac{3}{30}&0&1\end{bmatrix}=\begin{bmatrix}1&0&0&0\\2&1&0&0\\\frac{1}{2}&\frac{3}{30}&1&0\\4&-\frac{4}{15}&0&1\end{bmatrix}$



now apply G.E.:

$$\begin{bmatrix} 2 & 3 & 2 & 1 \\ 0 & -15 & -2 & -1 \\ 0 & 0 & \frac{12}{10} & \frac{6}{10} \\ 0 & 0 & -\frac{38}{15} & \frac{11}{15} \end{bmatrix} \xrightarrow{L_{43}=\frac{-19}{9}} \begin{bmatrix} 2 & 3 & 2 & 1 \\ 0 & -15 & -2 & -1 \\ 0 & 0 & \frac{12}{10} & \frac{6}{10} \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Finished.. hence

$$LU = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{10} & 1 & 0 \\ 4 & \frac{-4}{15} & \frac{-19}{9} & 1 \end{bmatrix} \begin{bmatrix} 2 & 3 & 2 & 1 \\ 0 & -15 & -2 & -1 \\ 0 & 0 & \frac{12}{10} & \frac{6}{10} \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

$PA = LU$.

$Ax = b \Rightarrow PAx = Pb \Rightarrow \boxed{LUx = Pb}$

Let $Ux = v$. then $Lv = Pb$. solve for v:

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{10} & 1 & 0 \\ 4 & \frac{-4}{15} & \frac{-19}{9} & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\nearrow P} \begin{bmatrix} 2 \\ 14 \\ -3 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 14 \\ 2 \\ -3 \end{bmatrix}$$

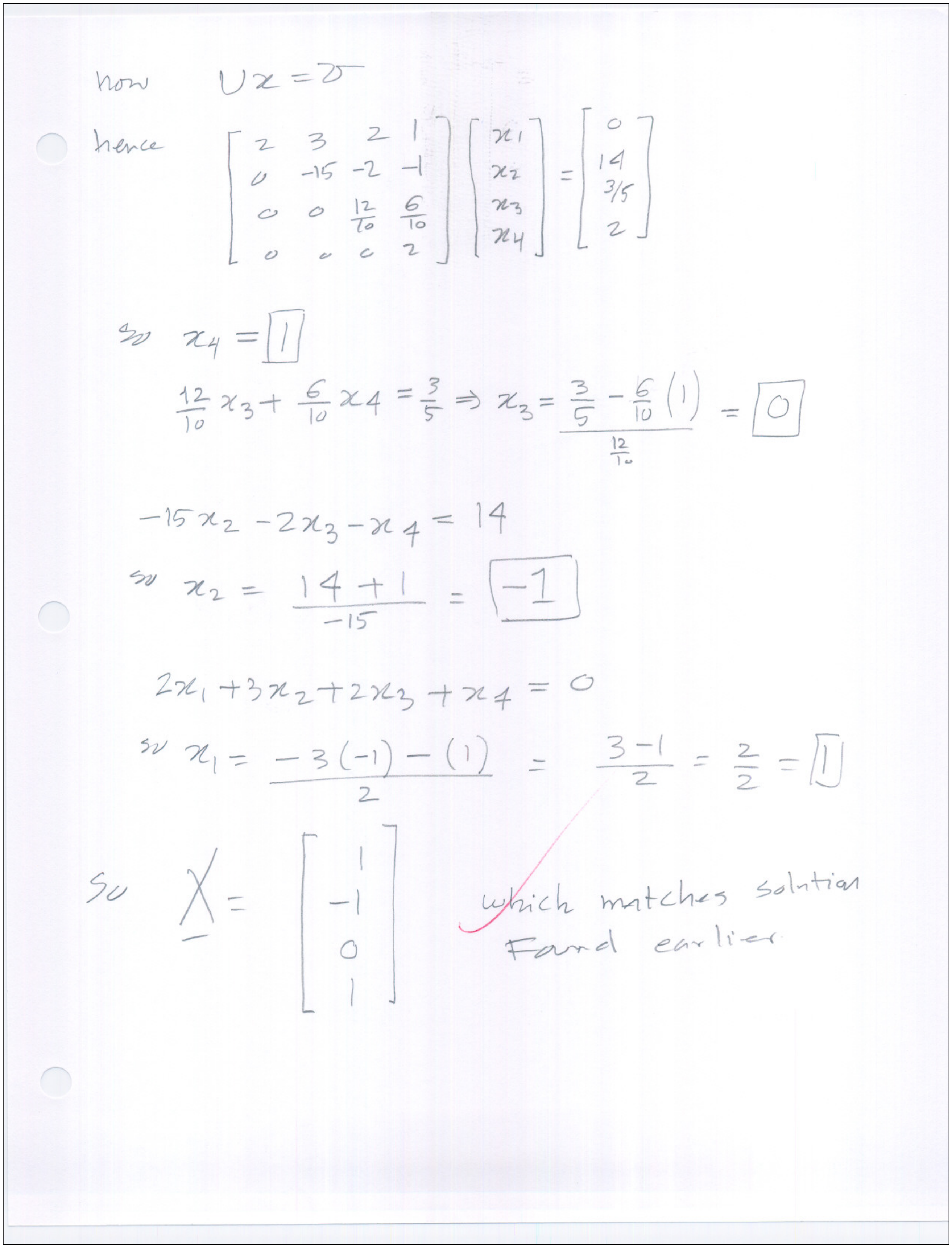
so $\boxed{v_1 = 0}$, $\boxed{v_2 = 14}$, $\frac{1}{10}v_2 + v_3 = 2 \Rightarrow v_3 = 2 - \frac{1}{10}(14) = 2 - \frac{14}{10}$

$v_3 = \frac{20-14}{10} = \frac{6}{10} = \boxed{\frac{3}{5}}$

$\frac{-4}{15}v_2 - \frac{19}{9}v_3 + v_4 = -3 \Rightarrow v_4 = -3 + \frac{4}{15}(14) + \frac{19}{9}\left(\frac{3}{5}\right) = -3 + \frac{56}{15} + \frac{19}{15}$

so $v_4 = \frac{-45+56+19}{15} = \frac{30}{15} = \boxed{2}$

$$\Rightarrow v = \begin{bmatrix} 0 \\ 14 \\ \frac{3}{5} \\ 2 \end{bmatrix} \longrightarrow$$



now $Ux = v$

hence
$$\begin{bmatrix} 2 & 3 & 2 & 1 \\ 0 & -15 & -2 & -1 \\ 0 & 0 & \frac{12}{10} & \frac{6}{10} \\ 0 & 0 & 0 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 14 \\ 3/5 \\ 2 \end{bmatrix}$$

so $x_4 = \boxed{1}$

$$\frac{12}{10}x_3 + \frac{6}{10}x4 = \frac{3}{5} \Rightarrow x_3 = \frac{\frac{3}{5} - \frac{6}{10}(1)}{\frac{12}{10}} = \boxed{0}$$

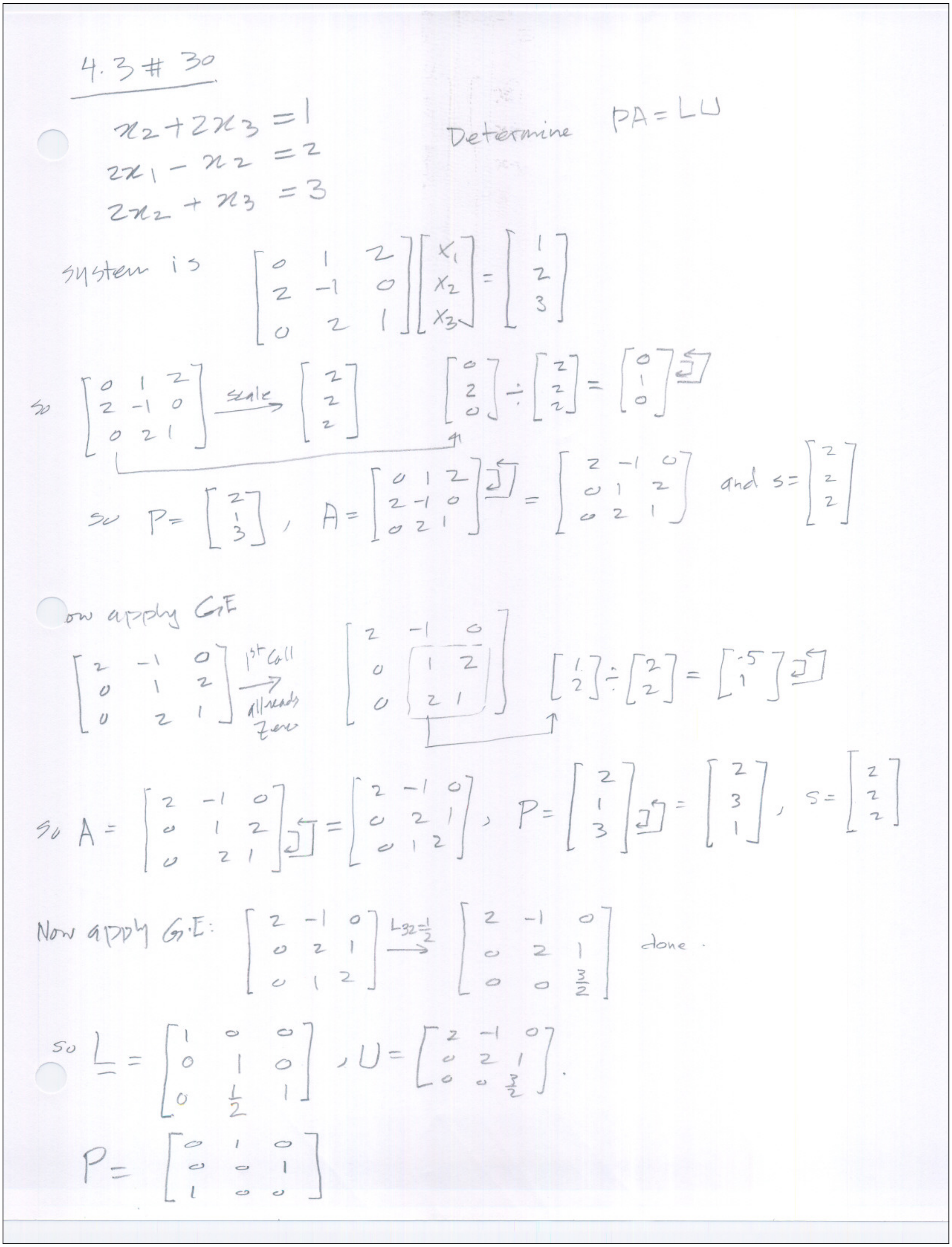
$$-15x_2 - 2x_3 - x_4 = 14$$

so $x_2 = \dfrac{14+1}{-15} = \boxed{-1}$

$$2x_1 + 3x_2 + 2x_3 + x_4 = 0$$

so $x_1 = \dfrac{-3(-1)-(1)}{2} = \dfrac{3-1}{2} = \dfrac{2}{2} = \boxed{1}$

So $\underline{X} = \begin{bmatrix} 1 \\ -1 \\ 0 \\ 1 \end{bmatrix}$ which matches solution Found earlier.



4.3 # 30

$$x_2 + 2x_3 = 1$$
$$2x_1 - x_2 = 2$$
$$2x_2 + x_3 = 3$$

Determine $PA = LU$

system is $\begin{bmatrix} 0 & 1 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$

so $\begin{bmatrix} 0 & 1 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow{\text{scale}} \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$ $\quad \begin{bmatrix} 0 \\ 2 \\ 0 \end{bmatrix} \div \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}^{\circlearrowleft}$

so $P = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}$, $A = \begin{bmatrix} 0 & 1 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix}^{\circlearrowleft} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}$ and $s = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$

Now apply GE

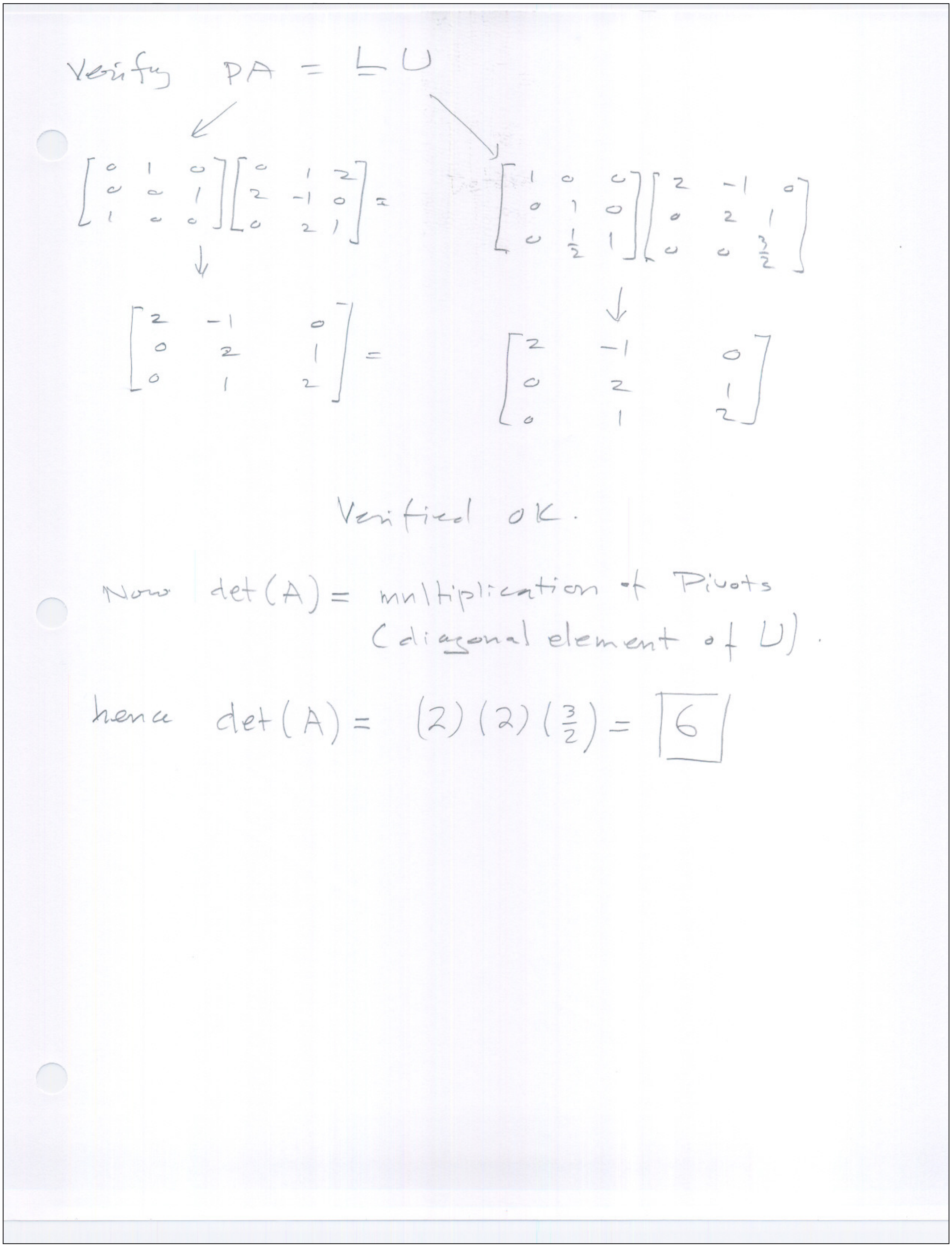
$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix} \xrightarrow[\text{allready Zero}]{1^{st}\text{ Coll}} \begin{bmatrix} 2 & -1 & 0 \\ 0 & \boxed{1 & 2} \\ 0 & 2 & 1 \end{bmatrix}$ $\quad \begin{bmatrix} 1 \\ 2 \end{bmatrix} \div \begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} .5 \\ 1 \end{bmatrix}^{\circlearrowleft}$

so $A = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 2 \\ 0 & 2 & 1 \end{bmatrix}_{\circlearrowleft} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$, $P = \begin{bmatrix} 2 \\ 1 \\ 3 \end{bmatrix}_{\circlearrowleft} = \begin{bmatrix} 2 \\ 3 \\ 1 \end{bmatrix}$, $s = \begin{bmatrix} 2 \\ 2 \\ 2 \end{bmatrix}$

Now apply G.E: $\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} \xrightarrow{L_{32}=\frac{1}{2}} \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}$ done.

so $L = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}$, $U = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}$.

$P = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$



Verify $PA = LU$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 & 2 \\ 2 & -1 & 0 \\ 0 & 2 & 1 \end{bmatrix} = \qquad \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{1}{2} & 1 \end{bmatrix}\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & \frac{3}{2} \end{bmatrix}$$

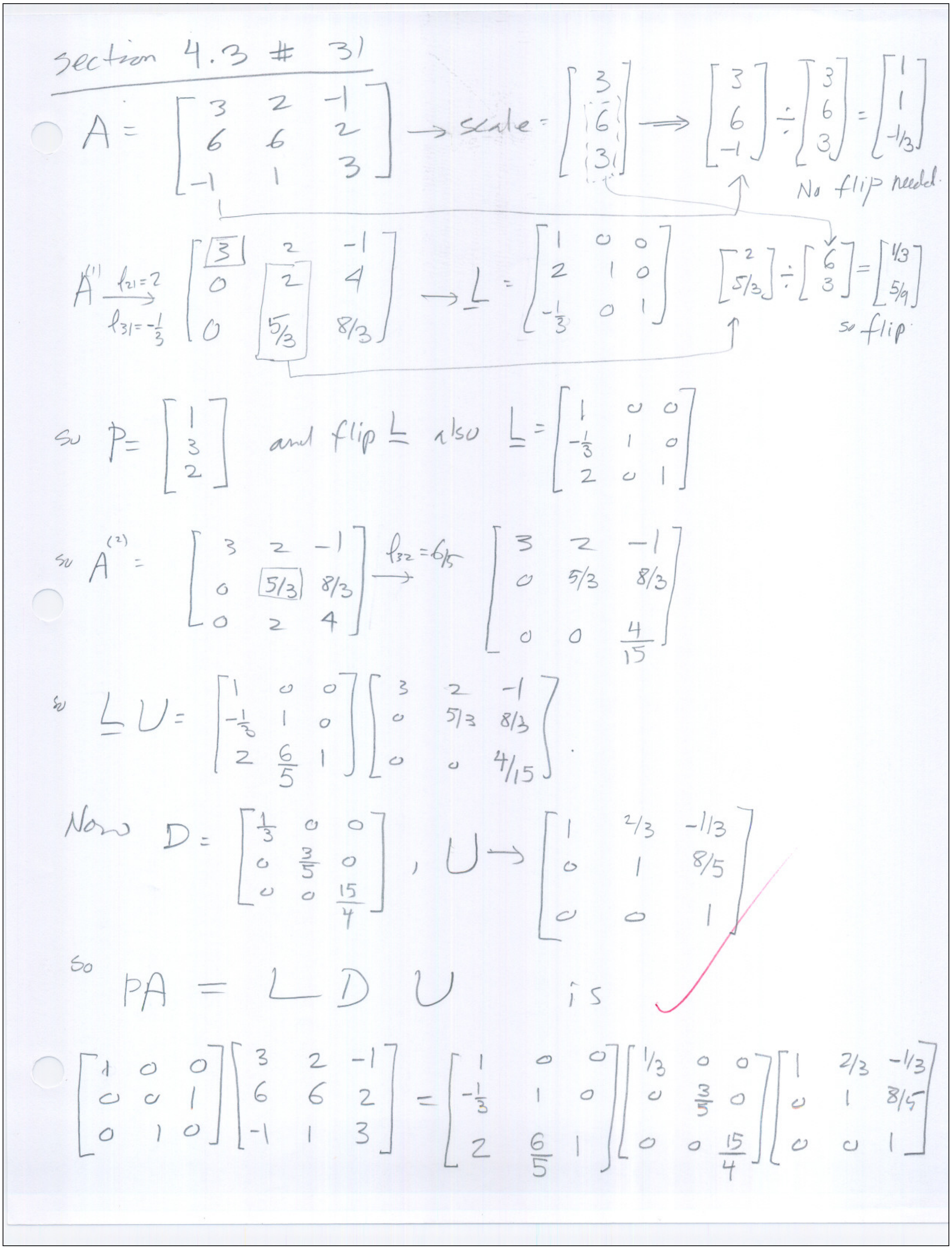
$$\downarrow \qquad\qquad\qquad\qquad \downarrow$$

$$\begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix} = \qquad \begin{bmatrix} 2 & -1 & 0 \\ 0 & 2 & 1 \\ 0 & 1 & 2 \end{bmatrix}$$

Verified OK.

Now $\det(A) =$ multiplication of Pivots (diagonal element of U).

hence $\det(A) = (2)(2)\left(\frac{3}{2}\right) = \boxed{6}$



section 4.3 # 31

$$A = \begin{bmatrix} 3 & 2 & -1 \\ 6 & 6 & 2 \\ -1 & 1 & 3 \end{bmatrix} \longrightarrow \text{scale} = \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix} \longrightarrow \begin{bmatrix} 3 \\ 6 \\ -1 \end{bmatrix} \div \begin{bmatrix} 3 \\ 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ -1/3 \end{bmatrix}$$

No flip needed.

$$A^{(1)} \xrightarrow[\ell_{31} = -\frac{1}{3}]{\ell_{21} = 2} \begin{bmatrix} \boxed{3} & 2 & -1 \\ 0 & \boxed{2} & 4 \\ 0 & \boxed{5/3} & 8/3 \end{bmatrix} \longrightarrow L = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \end{bmatrix} \qquad \begin{bmatrix} 2 \\ 5/3 \end{bmatrix} \div \begin{bmatrix} 6 \\ 3 \end{bmatrix} = \begin{bmatrix} 1/3 \\ 5/9 \end{bmatrix}$$

so flip

$$\text{so } P = \begin{bmatrix} 1 \\ 3 \\ 2 \end{bmatrix} \text{ and flip } L \text{ also } L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{3} & 1 & 0 \\ 2 & 0 & 1 \end{bmatrix}$$

$$\text{so } A^{(2)} = \begin{bmatrix} 3 & 2 & -1 \\ 0 & \boxed{5/3} & 8/3 \\ 0 & 2 & 4 \end{bmatrix} \xrightarrow{\ell_{32} = 6/5} \begin{bmatrix} 3 & 2 & -1 \\ 0 & 5/3 & 8/3 \\ 0 & 0 & \frac{4}{15} \end{bmatrix}$$

$$\text{so } LU = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{3} & 1 & 0 \\ 2 & \frac{6}{5} & 1 \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 0 & 5/3 & 8/3 \\ 0 & 0 & 4/15 \end{bmatrix}.$$

$$\text{Now } D = \begin{bmatrix} \frac{1}{3} & 0 & 0 \\ 0 & \frac{3}{5} & 0 \\ 0 & 0 & \frac{15}{4} \end{bmatrix}, \quad U \to \begin{bmatrix} 1 & 2/3 & -1/3 \\ 0 & 1 & 8/5 \\ 0 & 0 & 1 \end{bmatrix}$$

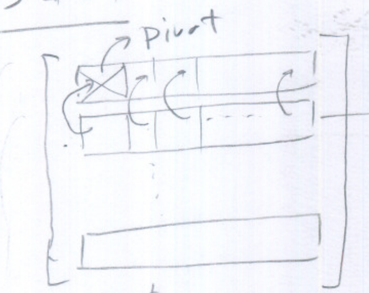
So

$$PA = LDU \text{ is}$$

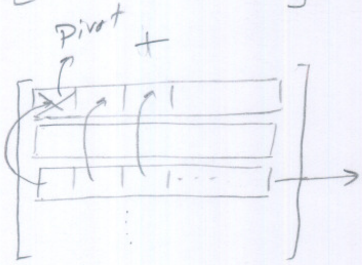
$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 & 2 & -1 \\ 6 & 6 & 2 \\ -1 & 1 & 3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{3} & 1 & 0 \\ 2 & \frac{6}{5} & 1 \end{bmatrix} \begin{bmatrix} 1/3 & 0 & 0 \\ 0 & \frac{3}{5} & 0 \\ 0 & 0 & \frac{15}{4} \end{bmatrix} \begin{bmatrix} 1 & 2/3 & -1/3 \\ 0 & 1 & 8/5 \\ 0 & 0 & 1 \end{bmatrix}$$

Section 4.3 # 39

given $A =$



$\to$ n multiplications by multiplier. l_{21} where $l_{21} = \frac{a_{21}}{a_{11}}$ (modulus sign).



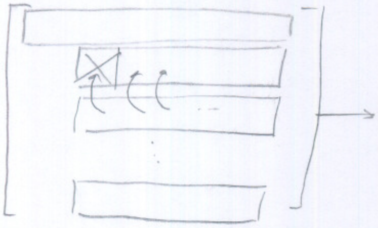
$\to$ n multiplications by multiplier l_{31} where $l_{31} = \frac{a_{31}}{a_{11}}$

$\vdots$

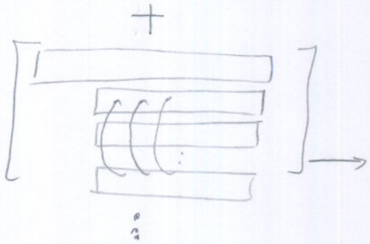
do the above for $(n-1)$ rows. $\Rightarrow$ $n(n-1)$ multiplications $+ (n-1)$ divisions $\left(\frac{a_{i1}}{a_{11}}\right)_{i=2..n}$

Now we repeate the above for $(n-1)\times(n-1)$ Matrix

$(n-2)$ times



$\to (n-1)$ ops



$\to (n-1)$ ops.

$\Rightarrow (n-1)(n-2)$ multiplications $+ (n-2)$ divisions

Continue this way we obtain

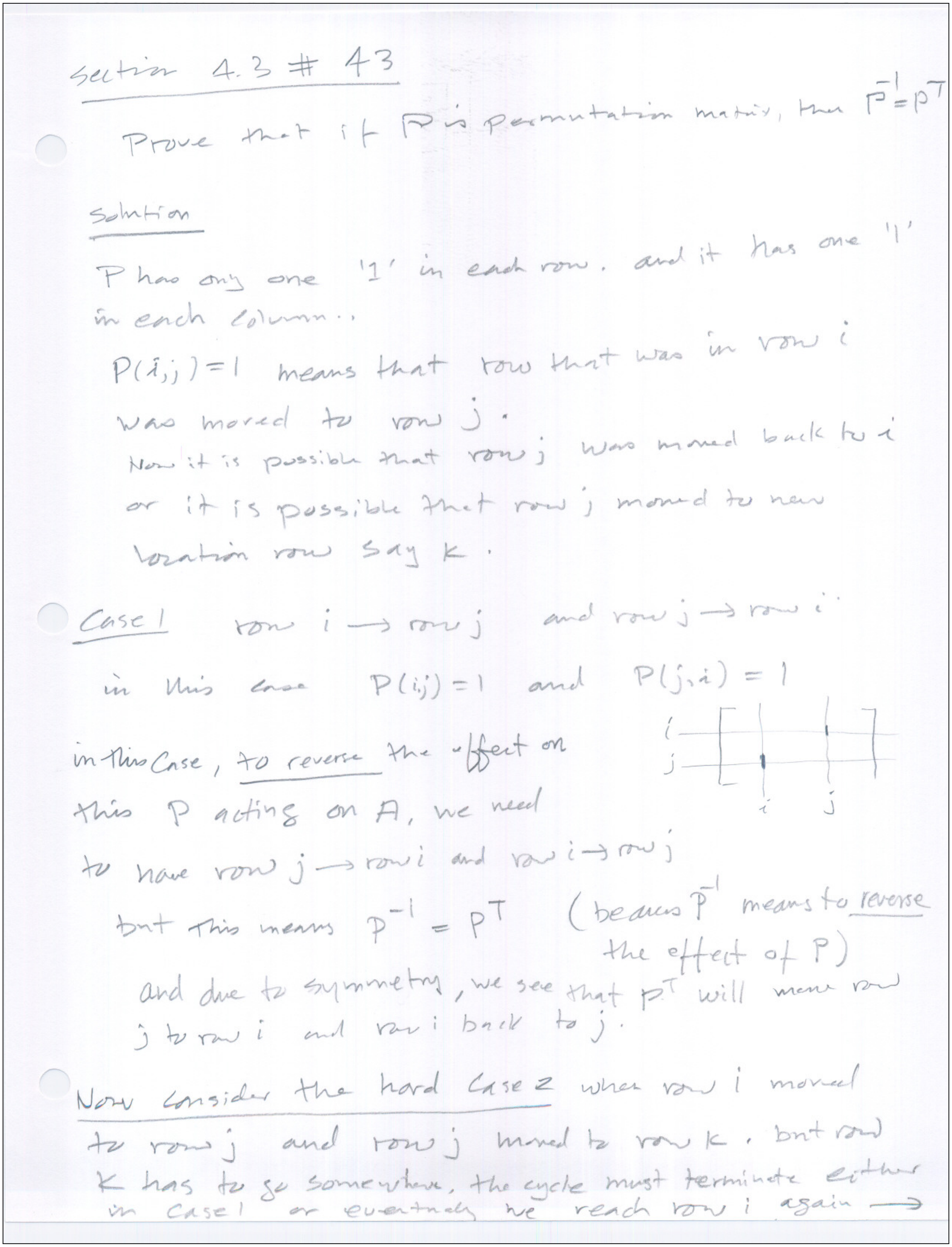
multiplications: $\overset{?}{n}(n-1) + (n-1)(n-2) + \cdots + 3\cdot 2 + 2\cdot 1 < n^2 + (n-1)^2 + \cdots + 3^2 + 2^2$ ✓

$= \frac{n}{6}(n+1)(2n+1) - 1$

$\simeq \boxed{\frac{n^3}{3}}$

divisions: $(n-1) + (n-2) + (n-3) + \cdots + 1$

$= \frac{n}{2}(n-1) \simeq \boxed{\frac{1}{2}n^2} \Rightarrow \text{ops} \cong \boxed{\frac{1}{3}n^3 + \frac{1}{2}n^2}$



Section 4.3 # 43

Prove that if P is permutation matrix, then $P^{-1}=P^T$

Solution

P has only one '1' in each row. and it has one '1' in each column..

$P(i,j)=1$ means that row that was in row i was moved to row j.

Now it is possible that row j was moved back to i or it is possible that row j moved to new location row say k.

Case 1 row $i \longrightarrow$ row j and row $j \longrightarrow$ row i

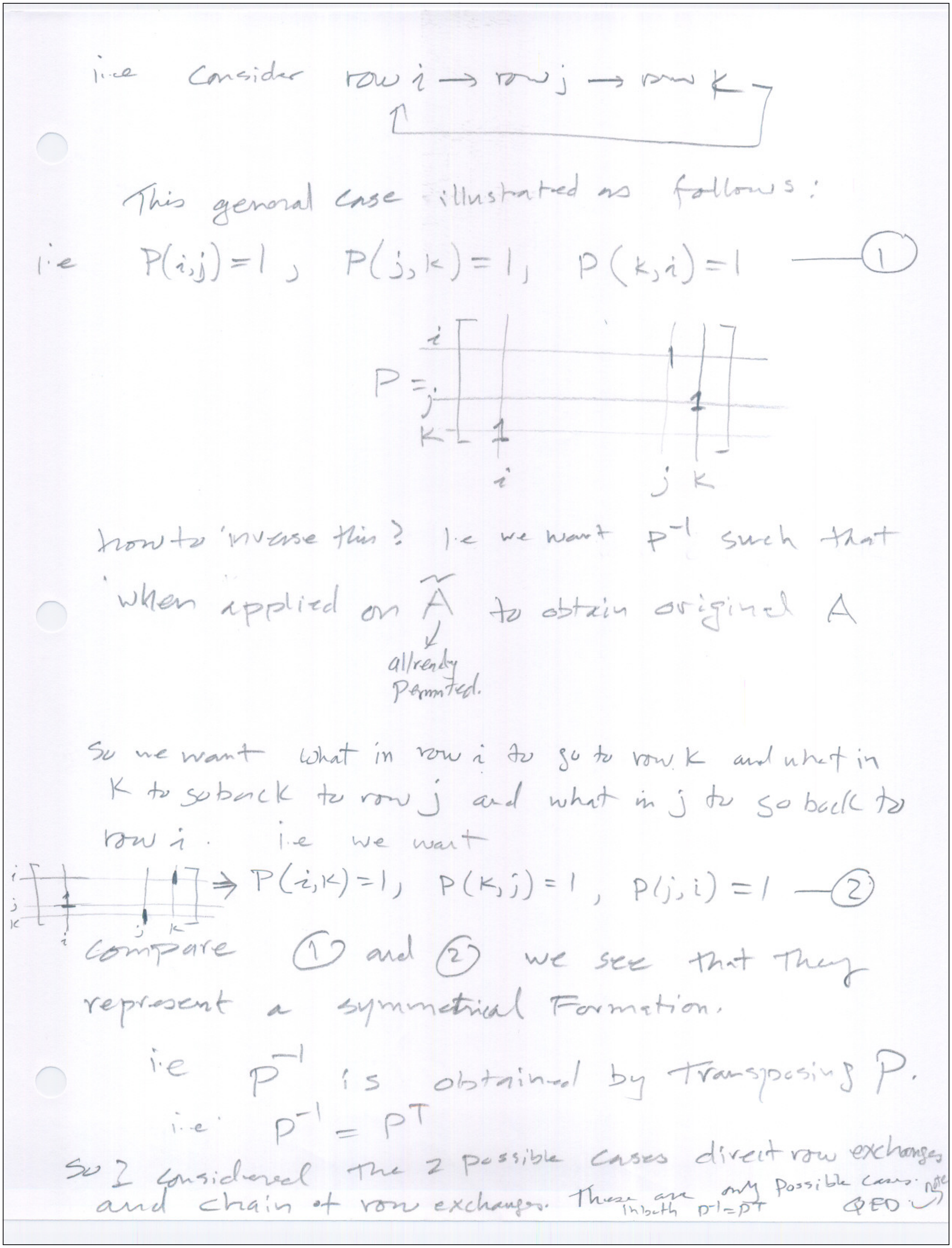
in this case $P(i,j)=1$ and $P(j,i)=1$

in this Case, to reverse the effect on this P acting on A, we need to have row $j \longrightarrow$ row i and row $i \longrightarrow$ row j

but this means $P^{-1}=P^T$ (because P^{-1} means to reverse the effect of P)

and due to symmetry, we see that P^T will move row j to row i and row i back to j.

Now consider the hard Case 2 when row i moved to row j and row j moved to row k. but row k has to go somewhere, the cycle must terminate either in Case 1 or eventually we reach row i again $\longrightarrow$



i.e Consider

$$\begin{tikzcd}\text{row } i \arrow[r] & \text{row } j \arrow[r] & \text{row } k \arrow[ll, bend left]\end{tikzcd}$$

This general case illustrated as follows:

i.e $P(i,j)=1$, $P(j,k)=1$, $P(k,i)=1$ —— ①

$$P = \begin{array}{c|ccc} & i & j & k \\ \hline i & & 1 & \\ j & & & 1 \\ k & 1 & & \end{array}$$

how to inverse this? i.e we want P^{-1} such that when applied on $\tilde{A}$ to obtain original A

$\tilde{A}$ ↓ allready Permuted.

So we want what in row i to go to row k and what in k to go back to row j and what in j to go back to row i. i.e we want

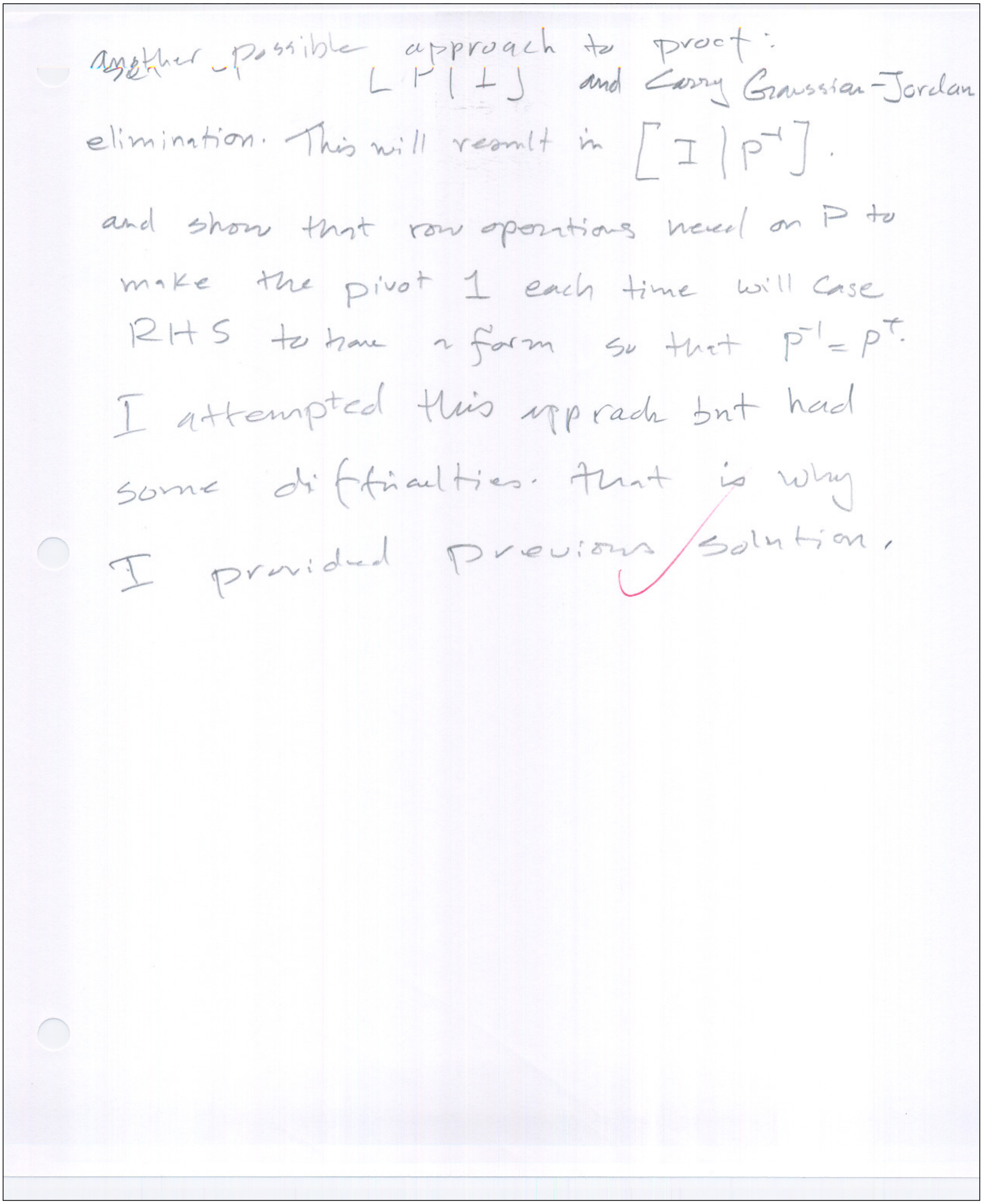
$$\begin{array}{c|ccc} & i & j & k \\ \hline i & & & 1 \\ j & 1 & & \\ k & & 1 & \end{array} \Rightarrow P(i,k)=1,\ P(k,j)=1,\ P(j,i)=1 \text{ —— ②}$$

compare ① and ② we see that they represent a symmetrical Formation.

i.e P^{-1} is obtained by Transposing P.

i.e $P^{-1}=P^T$

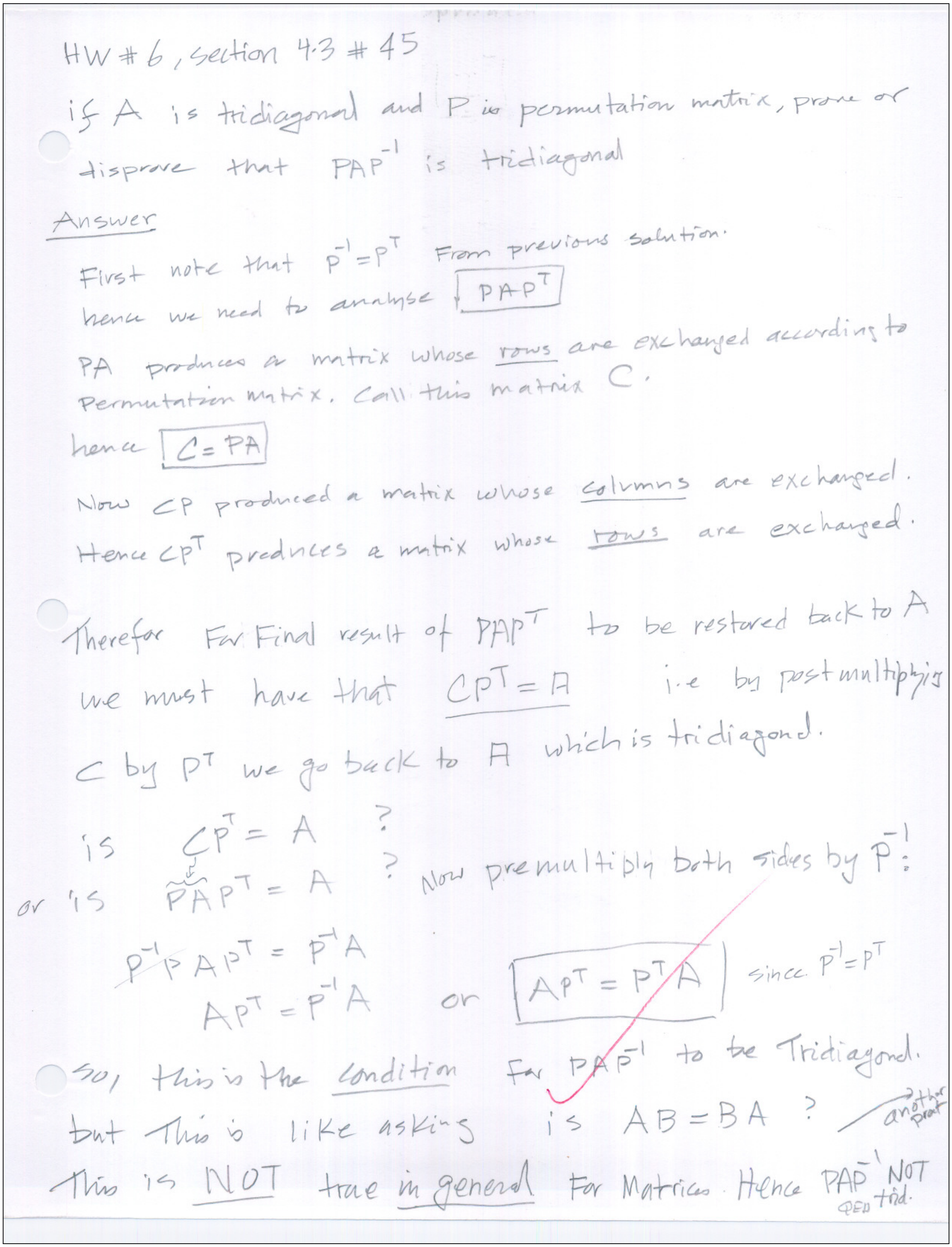
So I considered the 2 possible cases direct row exchanges and chain of row exchanges. These are only possible cases. note in both $P^{-1}=P^T$ QED



another possible approach to proof: $[P \mid I]$ and Carry Gaussian-Jordan elimination. This will result in $[I \mid P^{-1}]$.

and show that row operations need on P to make the pivot 1 each time will case RHS to have a form so that $P^{-1} = P^{T}$.

I attempted this approach but had some difficulties. that is why I provided previous solution.



HW # 6, section 4.3 # 45

if A is tridiagonal and P is permutation matrix, prove or disprove that PAP^{-1} is tridiagonal

Answer

First note that $P^{-1}=P^T$ From previous solution.

hence we need to analyse $\boxed{PAP^T}$

PA produces a matrix whose rows are exchanged according to permutation matrix. Call this matrix C.

hence $\boxed{C=PA}$

Now CP produced a matrix whose columns are exchanged.

Hence CP^T produces a matrix whose rows are exchanged.

Therefor For Final result of PAP^T to be restored back to A we must have that $CP^T=A$ i.e by post multiplying C by P^T we go back to A which is tridiagonal.

is $CP^T = A$?

or is $\underbrace{PA}P^T = A$? Now premultiply both sides by P^{-1}:

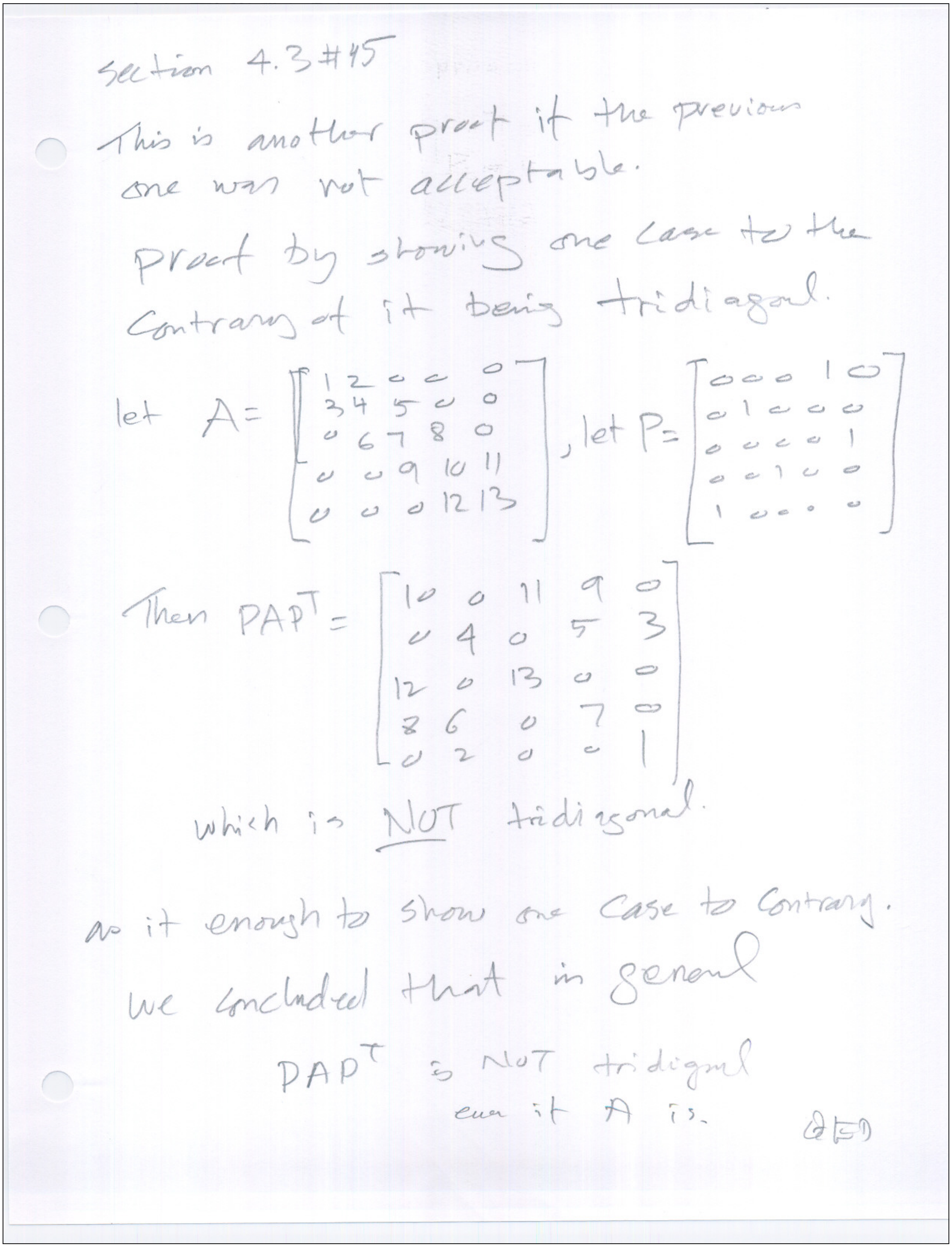
$$\cancel{P^{-1}P}AP^T = P^{-1}A$$

$$AP^T = P^{-1}A \quad \text{or} \quad \boxed{AP^T = P^TA} \quad \text{since } P^{-1}=P^T$$

So, this is the condition For PAP^{-1} to be Tridiagonal.

but This is like asking is $AB = BA$? another proof

This is NOT true in general For Matrices. Hence PAP^{-1} NOT trid. QED



section 4.3 #45

This is another proof if the previous one was not acceptable.

proof by showing one case to the contrary of it being tridiagonal.

let $A = \begin{bmatrix} 1 & 2 & 0 & 0 & 0 \\ 3 & 4 & 5 & 0 & 0 \\ 0 & 6 & 7 & 8 & 0 \\ 0 & 0 & 9 & 10 & 11 \\ 0 & 0 & 0 & 12 & 13 \end{bmatrix}$, let $P = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{bmatrix}$

Then $PAP^T = \begin{bmatrix} 10 & 0 & 11 & 9 & 0 \\ 0 & 4 & 0 & 5 & 3 \\ 12 & 0 & 13 & 0 & 0 \\ 8 & 6 & 0 & 7 & 0 \\ 0 & 2 & 0 & 0 & 1 \end{bmatrix}$

which is NOT tridiagonal.

as it enough to show one case to contrary.

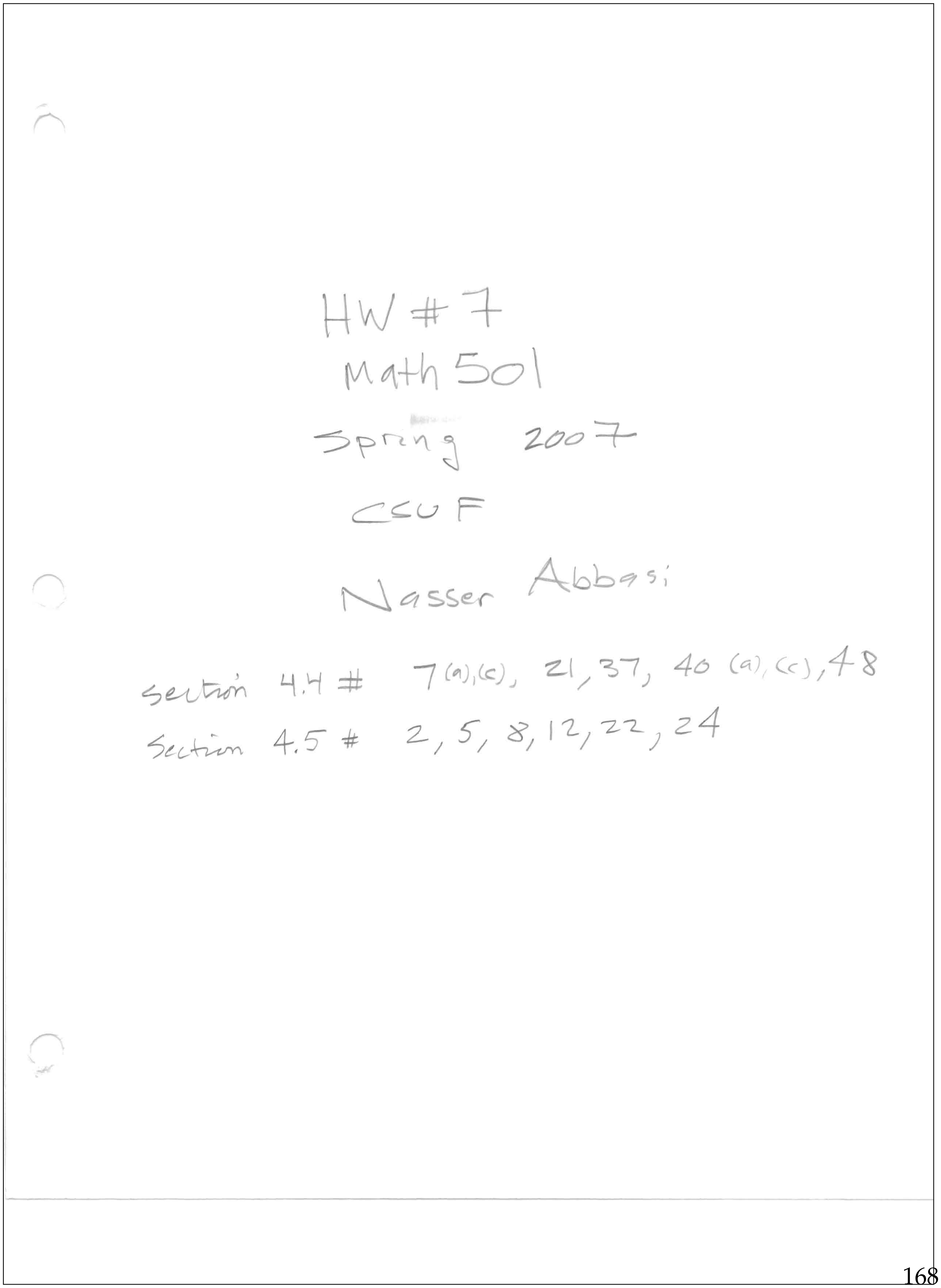
we concluded that in general

PAP^T is NOT tridiagonal

even if A is.

QED

4.8 HW 7



HW # 7

Math 501

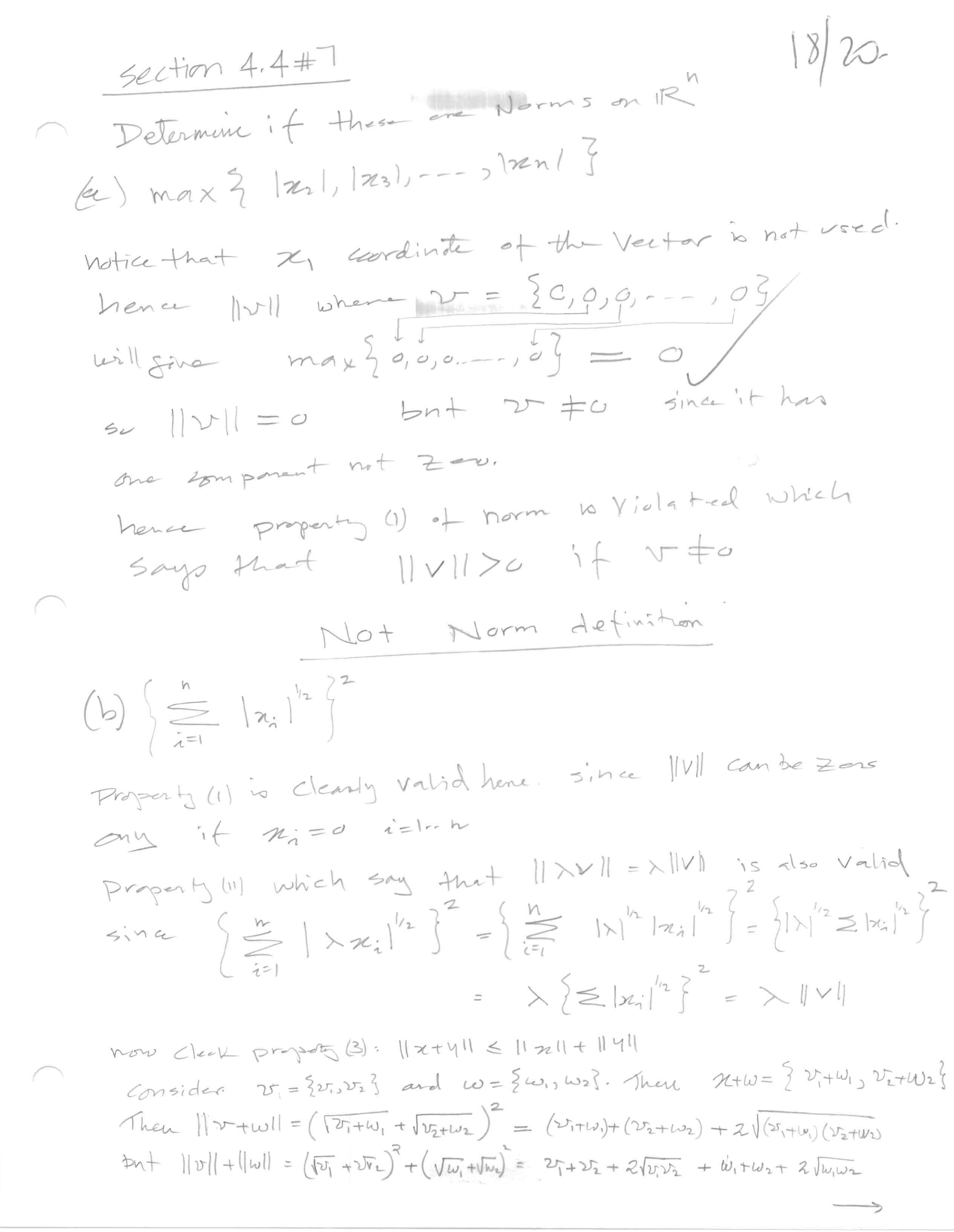
Spring 2007

CSUF

Nasser Abbasi

section 4.4 # 7(a),(c), 21, 37, 40 (a),(c), 48

Section 4.5 # 2, 5, 8, 12, 22, 24



section 4.4 #7

18/20

Determine if these are Norms on $\mathbb{R}^n$

(a) $\max\{|x_2|, |x_3|, ---, |x_n|\}$

notice that x_1 cordinate of the Vector is not used.

hence $||v||$ where $v = \{c, 0, 0, ---, 0\}$

will give $\max\{0, 0, 0 ---, 0\} = 0$ ✓

so $||v|| = 0$ but $v \neq 0$ since it has

one component not zero.

hence property (1) of norm is violated which says that $||v|| > 0$ if $v \neq 0$

Not Norm definition

(b) $\left\{\sum_{i=1}^{n} |x_i|^{1/2}\right\}^2$

Property (1) is clearly valid here. since $||v||$ can be zero only if $x_i = 0$ $i = 1 .. n$

Property (ii) which say that $||\lambda v|| = \lambda ||v||$ is also valid

since $\left\{\sum_{i=1}^{n} |\lambda x_i|^{1/2}\right\}^2 = \left\{\sum_{i=1}^{n} |\lambda|^{1/2} |x_i|^{1/2}\right\}^2 = \left\{|\lambda|^{1/2} \sum |x_i|^{1/2}\right\}^2$

$= \lambda \left\{\sum |x_i|^{1/2}\right\}^2 = \lambda ||v||$

now check property (3): $||x+y|| \leq ||x|| + ||y||$

consider $v_1 = \{v_1, v_2\}$ and $w = \{w_1, w_2\}$. Then $x + w = \{v_1 + w_1, v_2 + w_2\}$

Then $||v + w|| = \left(\sqrt{v_1 + w_1} + \sqrt{v_2 + w_2}\right)^2 = (v_1 + w_1) + (v_2 + w_2) + 2\sqrt{(v_1 + w_1)(v_2 + w_2)}$

but $||v|| + ||w|| = \left(\sqrt{v_1} + 2\sqrt{v_2}\right)^2 + \left(\sqrt{w_1} + \sqrt{w_2}\right)^2 = v_1 + v_2 + 2\sqrt{v_1 v_2} + w_1 + w_2 + 2\sqrt{w_1 w_2}$

$\longrightarrow$

so we ask

is $\|v+w\| \leq \|v\| + \|w\|$?

is $\left(\sqrt{v_1+w_1} + \sqrt{v_2+w_2}\right)^2 \leq \left(\sqrt{v_1}+\sqrt{v_2}\right)^2 + \left(\sqrt{w_1}+\sqrt{w_2}\right)^2$

~~try $v=(10,0)$, $w=(0,10)$~~.

~~$\|v+w\| = \left(\sqrt{v_1+w_1} + \sqrt{v_2+w_2}\right)^2 = \left(\sqrt{10}+\sqrt{10}\right)^2$~~

to make it simpler. try $v=(1,0)$, $w=(0,1)$

so $\|v+w\| = \left(\sqrt{v_1+w_1} + \sqrt{v_2+w_2}\right)^2 = (1+1)^2 = 4$

$$\|v\| + \|w\| = \left(\sqrt{v_1}+\sqrt{v_2}\right)^2 + \left(\sqrt{w_1}+\sqrt{w_2}\right)^2$$

$$= 1^2 + 1^2 = 2$$

so $\boxed{\|v+w\| > \|v\| + \|w\|}$

so property 3 NOT satisfied

so $\boxed{\text{NOT Norm}}$

Section 4.4 # 21

let $n=3$ and let $A=\begin{bmatrix} 4 & -3 & 2 \\ -1 & 0 & 5 \\ 2 & 6 & -2 \end{bmatrix}$

among all the vectors x satisfying $\|x\|_\infty \leq 1$, find one for which $\|Ax\|_\infty$ is as large as possible. Also give numerical value of $\|A\|_\infty$.

Answer

$$\|Ax\|_\infty \leq \|A\|_\infty \|x\|_\infty$$

but $\|A\|_\infty = \max \overset{\text{abs.}}{\nearrow} \text{sum of rows of } A$

$= \max \begin{bmatrix} 9 \\ 6 \\ 10 \end{bmatrix} = \boxed{10}$

so $\|Ax\|_\infty \leq 10 \|x\|_\infty$

so $\max \|Ax\|_\infty$ is when $\|x\|_\infty$ is max. which is 1

so $\|Ax\|_\infty \leq 10 \longrightarrow$ max is $\boxed{10}$ ✓

the x vector which satisfies this is when $\|x\|_\infty = 1$

so need a vector whose max coordinate is 1

so $\cancel{x = \begin{Bmatrix} 1 \\ 0 \\ 0 \end{Bmatrix}}$ will do.

$Ax = \begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}$

$\|Ax\|_\infty = \left\|\begin{bmatrix} 4 \\ -1 \\ 2 \end{bmatrix}\right\|_\infty = 4$ not 10.

section 4.5 #37

Prove these properties

(a) $\|0\|=0$

(b) $\|x+y\| \geq | \|x\| - \|y\| |$

(c) $\| \sum_{i=1}^{m} x^{(i)} \| \leq \sum_{i=1}^{m} \| x^{(i)} \|$ for vectors $x^{(1)}, x^{(2)}, \dots, x^{(m)}$

(a) from properties of Norms, we know that

$$\|\lambda x\| = \lambda \|x\| \quad \text{for } \lambda \text{ any constant}$$

so let $x=0$, hence we write

$$\| \underbrace{\lambda 0} \| = \lambda \|0\|$$

but $\lambda(0) = 0$ since scalar multiplication.

so $\|0\| = \lambda \|0\|$

for this to be valid for any non zero λ, it must be that $\boxed{\|0\| = 0}$

This is like saying $a=3a \rightarrow$ only true if $a=0$

(b) $\|y\| = \|y+(x-x)\| = \|(y+x)+(-x)\|$ ———

but since $\|A+B\| \leq \|A\|+\|B\|$, then above is

$\leq \|(y+x)\| + \|(-x)\|$ ←

so $\|y\| \leq \|(y+x)\| + \|(-x)\|$

so $\|y\| \leq \|y+x\| + \|x\|$ since $\|-x\| = \|x\|$

so $\boxed{\|y\|-\|x\| \leq \|y+x\|}$ by moving $\|x\|$ to LHS.

QED

→

(c) prove

$$\left\| \sum_{i=1}^{m} x^{(i)} \right\| \leq \sum_{i=1}^{m} \| x^{(i)} \| \qquad \text{For vectors } x^{(1)}, x^{(2)}, \ldots, x^{(m)}.$$

LHS is $\| x^1 + x^2 + \quad + x^m \| = \| x^1 + (x^2 + \cdots + x^m) \|$

$\leq \| x^1 \| + \| x^2 + x^3 + \cdots + x^m \|$

by triangle inequality.

Repeate the above on $\| x^2 + x^3 + \cdots + x^m \|$ we get

$$\| x^1 + x^2 + \cdots + x^m \| \leq \| x^{(1)} \| + \| x^{(2)} + (x^{(3)} + x^{(4)} + \cdots + x^{(m)}) \|$$

$$\leq \| x^{(1)} \| + \| x^{(2)} \| + \| x^{(3)} + x^{(4)} + \cdots + x^{(m)} \|$$

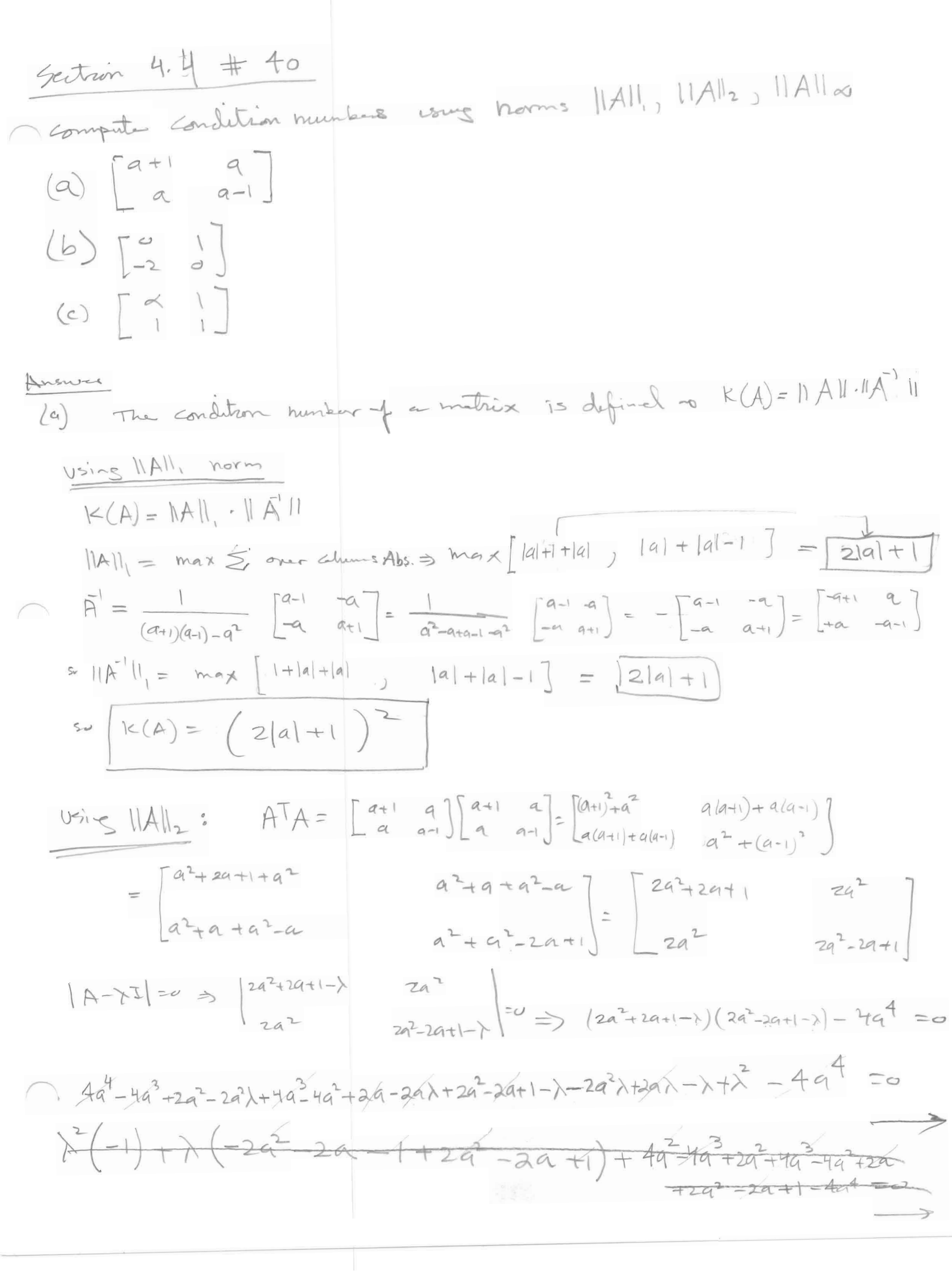
$\vdots$

etc

so we obtain

$$\| x^{(1)} + x^{(2)} + \cdots + x^{(m)} \| \leq \| x^{(1)} \| + \| x^{(2)} \| + \cdots + \| x^{(m)} \|$$

i.e $\boxed{\left\| \sum_{i=1}^{m} x^{(i)} \right\| \leq \sum_{i=1}^{m} \| x^{(i)} \|}$



Section 4.4 # 40

compute condition numbers using norms $\|A\|_1$, $\|A\|_2$, $\|A\|_\infty$

(a) $\begin{bmatrix} a+1 & a \\ a & a-1 \end{bmatrix}$

(b) $\begin{bmatrix} 0 & 1 \\ -2 & 0 \end{bmatrix}$

(c) $\begin{bmatrix} \alpha & 1 \\ 1 & 1 \end{bmatrix}$

Answer

(a) The condition number of a matrix is defined as $K(A)=\|A\|\cdot\|A^{-1}\|$

Using $\|A\|_1$ norm

$K(A)=\|A\|_1\cdot\|A^{-1}\|$

$\|A\|_1 =$ max $\sum$ over columns Abs. $\Rightarrow$ max$\left[|a|+1+|a|,\ |a|+|a|-1\right] = \boxed{2|a|+1}$

$A^{-1} = \frac{1}{(a+1)(a-1)-a^2}\begin{bmatrix} a-1 & -a \\ -a & a+1 \end{bmatrix} = \frac{1}{a^2-a+a-1-a^2}\begin{bmatrix} a-1 & -a \\ -a & a+1 \end{bmatrix} = -\begin{bmatrix} a-1 & -a \\ -a & a+1 \end{bmatrix} = \begin{bmatrix} -a+1 & a \\ +a & -a-1 \end{bmatrix}$

so $\|A^{-1}\|_1 =$ max$\left[1+|a|+|a|,\ |a|+|a|-1\right] = \boxed{2|a|+1}$

so $\boxed{K(A) = (2|a|+1)^2}$

Using $\|A\|_2$: $A^TA = \begin{bmatrix} a+1 & a \\ a & a-1 \end{bmatrix}\begin{bmatrix} a+1 & a \\ a & a-1 \end{bmatrix} = \begin{bmatrix} (a+1)^2+a^2 & a(a+1)+a(a-1) \\ a(a+1)+a(a-1) & a^2+(a-1)^2 \end{bmatrix}$

$= \begin{bmatrix} a^2+2a+1+a^2 & a^2+a+a^2-a \\ a^2+a+a^2-a & a^2+a^2-2a+1 \end{bmatrix} = \begin{bmatrix} 2a^2+2a+1 & 2a^2 \\ 2a^2 & 2a^2-2a+1 \end{bmatrix}$

$|A-\lambda I|=0 \Rightarrow \begin{vmatrix} 2a^2+2a+1-\lambda & 2a^2 \\ 2a^2 & 2a^2-2a+1-\lambda \end{vmatrix}=0 \Rightarrow (2a^2+2a+1-\lambda)(2a^2-2a+1-\lambda)-4a^4=0$

$4a^4-4a^3+2a^2-2a^2\lambda+4a^3-4a^2+2a-2a\lambda+2a^2-2a+1-\lambda-2a^2\lambda+2a\lambda-\lambda+\lambda^2-4a^4=0$ $\longrightarrow$

~~$\lambda^2(-1)+\lambda(-2a^2-2a-1+2a^2$~~ $-2a+1)+$ ~~$4a^2-4a^3+2a^2+4a^3-4a^2+2a$~~

~~$+2a^2-2a+1-4a^4=0$~~ $\longrightarrow$

$$2a^2-2a^2\lambda-4a^2+2a^2+1-\lambda=2a^2\lambda-\lambda+\lambda^2=0$$

$$\lambda^2+\lambda(-2a^2-1-2a^2-1)+4a^2-4a^2+1=0$$

$$\boxed{\lambda^2+\lambda(-4a^2-2)+1=0}$$

so $\lambda=\dfrac{-b\pm\sqrt{b^2-4ac}}{2a}=\dfrac{4a^2+2\pm\sqrt{(-4a^2-2)^2-4}}{2}$

$$=2a^2+1\pm\frac{1}{2}\sqrt{16a^4+4+16a^2-4}=2a^2+1\pm\frac{1}{2}\sqrt{16a^2(a^2+1)}$$

$$\boxed{\lambda=2a^2+1\pm 2a\sqrt{a^2+1}}$$

$\lambda_1=2a^2+1+2a\sqrt{a^2+1}$ $\qquad\lambda_2=2a^2+1-2a\sqrt{a^2+1}$

λ_1 is larger of the 2 eigenvalues

so $\|A\|_2=\sqrt{\lambda_1}=\sqrt{2a^2+1+2a\sqrt{a^2+1}}$

now to find $\|A^{-1}\|_2$:

$$(A^{-1})^T A^{-1}=\begin{bmatrix}-a+1 & a\\ a & -a-1\end{bmatrix}^T\begin{bmatrix}-a+1 & a\\ a & -a-1\end{bmatrix}=\begin{bmatrix}-a+1 & a\\ a & -a-1\end{bmatrix}\begin{bmatrix}-a+ & a\\ a & -a-1\end{bmatrix}$$

$$=\begin{bmatrix}a^2-a-a+1+a^2 & -a^2+a-a^2-a\\ -a^2+a-a^2-a & a^2+a^2+a+a+1\end{bmatrix}=\begin{bmatrix}2a^2-2a+1 & -2a^2\\ -2a^2 & 2a^2+2a+1\end{bmatrix}$$

$\downarrow$

compare this to A^TA before

$$\begin{bmatrix}2a^2+2a+1 & 2a^2\\ 2a^2 & 2a^2-2a+1\end{bmatrix}$$

negative signs on diagonal cancel out
we see it will come out the
same charateristic polynomial in λ.
$\Longrightarrow$ λmax as befor. $\Longrightarrow$

so $\|A\|_2 \; \|A^{-1}\|_2 = \sqrt{\lambda} \; \sqrt{\lambda} = \lambda$

$$\boxed{K(A) = 2a^2+1+2a\sqrt{a^2+1}}$$

Now do $\|A\|_\infty$

$K(A) = \|A\|_\infty \; \|A^{-1}\|_\infty$

but $\|A\|_\infty = \max\limits_{\text{rows}} \{ |a|+1+|a| , |a|+|a|-1 \} = 2|a|+1$

$$\|A^{-1}\|_\infty = \left\| \begin{bmatrix} -a+1 & a \\ a & -a-1 \end{bmatrix} \right\|_\infty = \max \{ |a|+1+|a| , |a|+|a|-1 \}$$

$$= \max \{ 2|a|+1 , 2|a|-1 \}$$

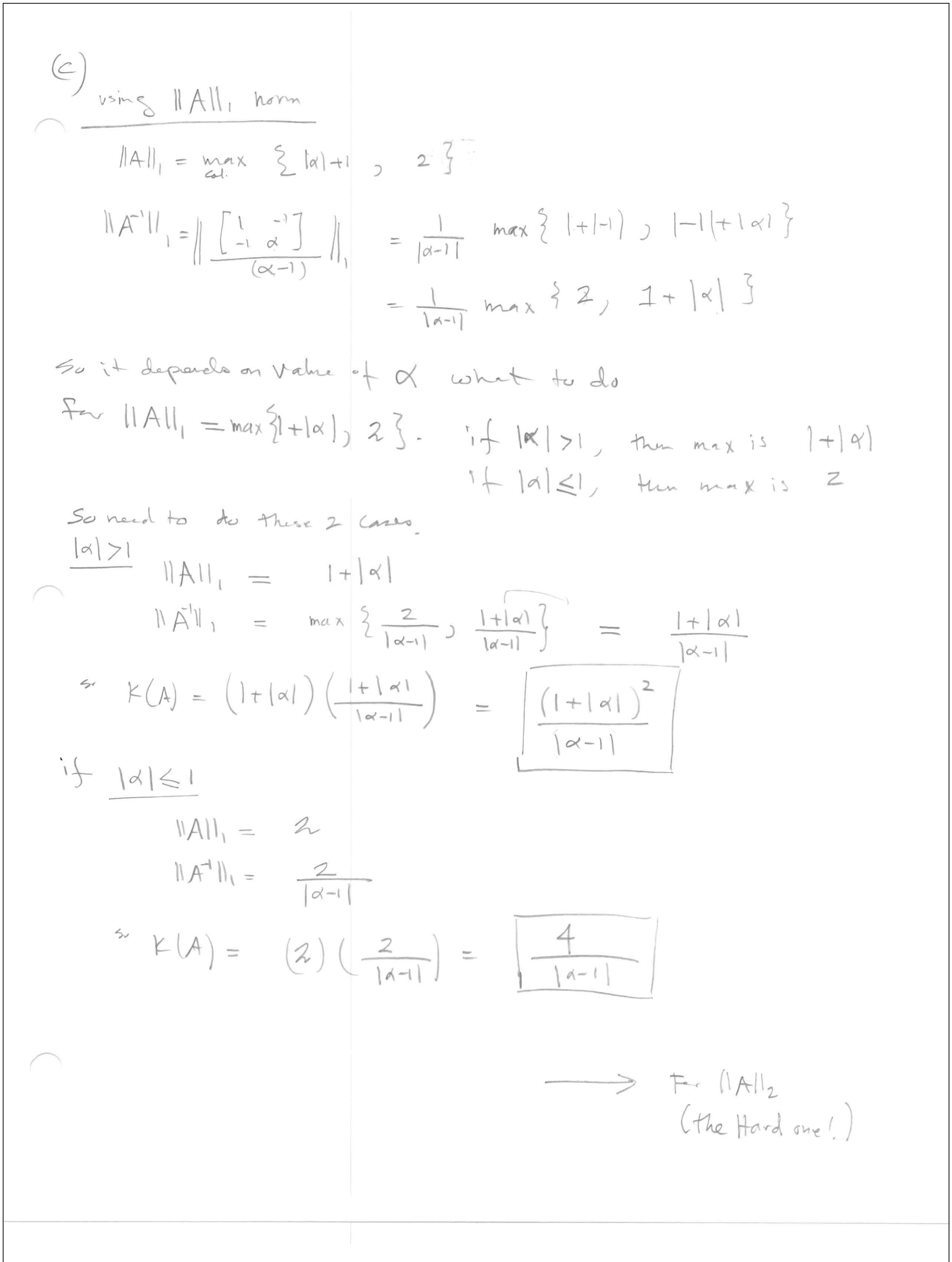
$$= 2|a|+1$$

so $K(A) = \|A\|_\infty \; \|A^{-1}\|_\infty$

$$\boxed{K(A) = (2|a|+1)^2}$$

Notice $\underset{\overset{\nearrow}{\text{norm}}}{K(A)_\infty} = \underset{\overset{\nearrow}{\text{norm.}}}{K(A)_1}$

$\longrightarrow$



(c)

using $\|A\|_1$ norm

$$\|A\|_1 = \max_{col.} \{ |\alpha|+1 , 2 \}$$

$$\|A^{-1}\|_1 = \left\| \frac{\begin{bmatrix} 1 & -1 \\ -1 & \alpha \end{bmatrix}}{(\alpha-1)} \right\|_1 = \frac{1}{|\alpha-1|} \max \{ |1|+|-1| , |-1|+|\alpha| \}$$

$$= \frac{1}{|\alpha-1|} \max \{ 2 , 1+|\alpha| \}$$

So it depends on value of α what to do

for $\|A\|_1 = \max\{1+|\alpha|, 2\}$. if $|\alpha|>1$, then max is $1+|\alpha|$

if $|\alpha| \leq 1$, then max is 2

So need to do these 2 cases.

$|\alpha|>1$

$$\|A\|_1 = 1+|\alpha|$$

$$\|A^{-1}\|_1 = \max \left\{ \frac{2}{|\alpha-1|} , \frac{1+|\alpha|}{|\alpha-1|} \right\} = \frac{1+|\alpha|}{|\alpha-1|}$$

so $K(A) = (1+|\alpha|) \left(\frac{1+|\alpha|}{|\alpha-1|} \right) = \boxed{\frac{(1+|\alpha|)^2}{|\alpha-1|}}$

if $|\alpha| \leq 1$

$$\|A\|_1 = 2$$

$$\|A^{-1}\|_1 = \frac{2}{|\alpha-1|}$$

so $K(A) = (2) \left(\frac{2}{|\alpha-1|} \right) = \boxed{\frac{4}{|\alpha-1|}}$

$\longrightarrow$ For $\|A\|_2$ (the Hard one!)

$\|A\|_2$

$$K(A) = \|A\|_2 \cdot \|A^{-1}\|_2$$

Find $\|A\|_2$: $A^TA = \begin{bmatrix}\alpha & 1\\ 1 & 1\end{bmatrix}^T \begin{bmatrix}\alpha & 1\\ 1 & 1\end{bmatrix} = \begin{bmatrix}\alpha & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\alpha & 1\\ 1 & 1\end{bmatrix} = \begin{bmatrix}\alpha^2+1 & \alpha+1\\ \alpha+1 & 2\end{bmatrix}$

Setup characteristic equation: $\begin{vmatrix}\alpha^2+1-\lambda & \alpha+1\\ \alpha+1 & 2-\lambda\end{vmatrix} = (\alpha^2+1-\lambda)(2-\lambda) - (\alpha+1)^2 = 0$

$$2\alpha^2 - \lambda\alpha^2 + 2 - \lambda - 2\lambda + \lambda^2 - \alpha^2 - 2\alpha - 1 = 0$$

$$\lambda^2 + \lambda(-\alpha^2 - 1 - 2) + (2\alpha^2 + 2 - \alpha^2 - 2\alpha - 1) = 0$$

$$\lambda^2 + \lambda(-\alpha^2 - 3) + (\alpha^2 - 2\alpha + 1) = 0$$

$$\lambda = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-(-\alpha^2-3)}{2} \pm \frac{1}{2}\sqrt{(-\alpha^2-3)^2 - 4(\alpha^2 - 2\alpha + 1)}$$

$$= \frac{\alpha^2+3}{2} \pm \frac{1}{2}\sqrt{\alpha^4 + 9 + 6\alpha^2 - 4\alpha^2 + 8\alpha - 4} = \frac{\alpha^2+3}{2} \pm \frac{1}{2}\sqrt{\alpha^4 + 2\alpha^2 + 8\alpha + 5}$$

so $\lambda_1 = \frac{\alpha^2+3}{2} + \frac{1}{2}\sqrt{\alpha^4 + 2\alpha^2 + 8\alpha + 5}$, $\lambda_2 = \frac{\alpha^2+3}{2} - \frac{1}{2}\sqrt{\alpha^4 + 2\alpha^2 + 8\alpha + 5}$

$\downarrow$

$\boxed{\lambda_{max} = \lambda_1}$ so $\sigma_1 = \sqrt{\lambda_1}$

so $\|A\|_2 = \sqrt{\lambda_1}$

now find $\|A^{-1}\|_2$: $A^{-1} = \frac{1}{|\alpha-1|}\begin{bmatrix}1 & -1\\ -1 & \alpha\end{bmatrix}$

so $(A^{-1})^T A^{-1} = \frac{1}{|\alpha-1|^2}\begin{bmatrix}1 & -1\\ -1 & \alpha\end{bmatrix}\begin{bmatrix}1 & -1\\ -1 & \alpha\end{bmatrix} = \frac{1}{|\alpha-1|^2}\begin{bmatrix}2 & -1-\alpha\\ -1-\alpha & 1+\alpha^2\end{bmatrix}$

Compare $\begin{bmatrix}\alpha^2+1 & \alpha+1\\ \alpha+1 & 2\end{bmatrix}$ with $\begin{bmatrix}2 & -(1+\alpha)\\ -(1+\alpha) & 1+\alpha^2\end{bmatrix}$ we see same λmax.

so λ_{max} For $(A^{-1})^T A^{-1}$ is $\frac{1}{|\alpha-1|^2}\lambda_1 \rightarrow$ found above

→

so $\|A^{-1}\|_2 = \sqrt{\lambda_{max}} = \sqrt{\frac{1}{|\alpha-1|^2}\lambda_1} = \frac{1}{|\alpha-1|}\sqrt{\lambda_1}$

so $K(A) = \|A\|_2 \cdot \|A^{-1}\|_2$

$$= \sqrt{\lambda_1}\,\frac{1}{|\alpha-1|}\sqrt{\lambda_1} = \boxed{\left(\frac{\lambda_1}{|\alpha-1|}\right)}$$

where $\lambda_1 = \frac{\alpha^2+3}{2} + \frac{1}{2}\sqrt{\alpha^4+2\alpha^2+8\alpha+5}$

Now do $\|A\|_\infty$

$\|A\|_\infty = \max_{row}\left\{ |\alpha|+1 , 2 \right\}$

$\|A^{-1}\|_\infty = \frac{1}{|\alpha-1|}\max\left\{ 2, 1+|\alpha| \right\}$

as befor for $\|A\|_1$; it depends on value for α

for $|\alpha|>1$, we set $\|A\|_\infty = 1+|\alpha|$

$\|A^{-1}\|_\infty = \frac{1+|\alpha|}{|\alpha-1|}$

so $K(A) = (1+|\alpha|)\left(\frac{1+|\alpha|}{|\alpha-1|}\right) = \boxed{\frac{(1+|\alpha|)^2}{|\alpha-1|}}$

for $|\alpha|\leq 1$

$\|A\|_\infty = 2$

$\|A^{-1}\|_\infty = \frac{2}{|\alpha-1|}$

so $K(A) = \boxed{\frac{4}{|\alpha-1|}}$

Section 4.5 # 48

Prove condition number has the property $\lambda \neq 0$

$$K(\lambda A) = K(A)$$

solution

$$K(A) \equiv \|A\| \cdot \|A^{-1}\|$$

so $K(\lambda A) = \|\lambda A\| \cdot \|(\lambda A)^{-1}\|$

but $(\lambda A)^{-1} = A^{-1}\lambda^{-1}$

so $K(\lambda A) = \|\lambda A\| \; \|A^{-1}\lambda^{-1}\|$

but $\|\lambda A\| = |\lambda| \; \|A\|$

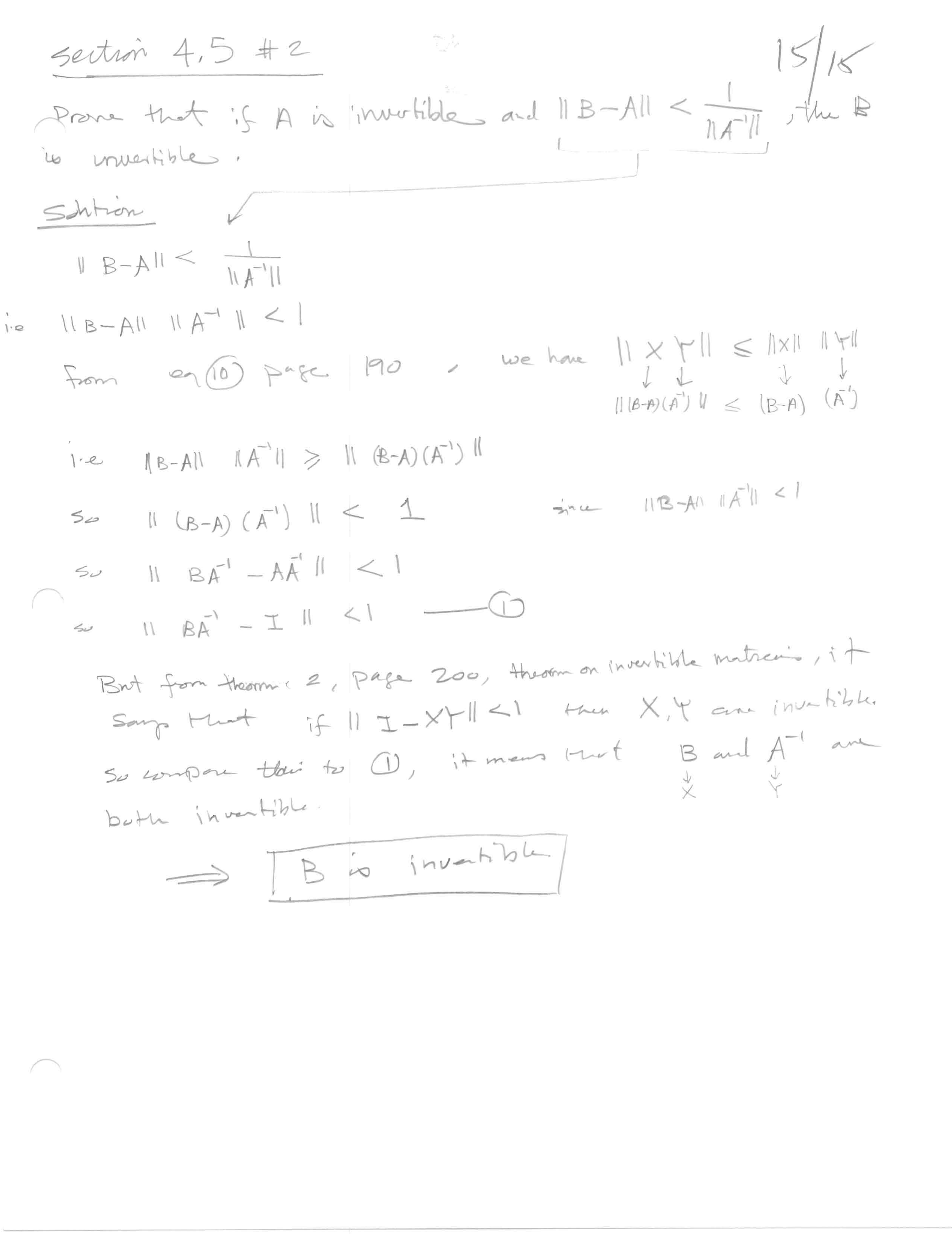
and $\|A^{-1}\lambda^{-1}\| = |\lambda^{-1}| \; \|A^{-1}\|$

so $K(\lambda A) = |\lambda| \; \|A\| \; |\lambda^{-1}| \; \|A^{-1}|$

$$K(\lambda A) = \underbrace{\|A\| \; \|A^{-1}\|}$$

$$K(\lambda A) = K(A)$$

QED



section 4.5 #2

15/15

Prove that if A is invertible and $\underbrace{\|B-A\| < \frac{1}{\|A^{-1}\|}}$, then B is invertible.

Solution

$\|B-A\| < \frac{1}{\|A^{-1}\|}$

i.e $\|B-A\|\,\|A^{-1}\| < 1$

from eq (10) page 190, we have $\|\underset{\overset{\downarrow}{(B-A)}}{X}\,\underset{\overset{\downarrow}{(A^{-1})}}{Y}\| \le \|\underset{\overset{\downarrow}{(B-A)}}{X}\|\,\|\underset{\overset{\downarrow}{(A^{-1})}}{Y}\|$

$\|(B-A)(A^{-1})\| \le (B-A)\ (A^{-1})$

i.e $\|B-A\|\,\|A^{-1}\| \geq \|(B-A)(A^{-1})\|$

so $\|(B-A)(A^{-1})\| < 1$ since $\|B-A\|\,\|A^{-1}\| < 1$

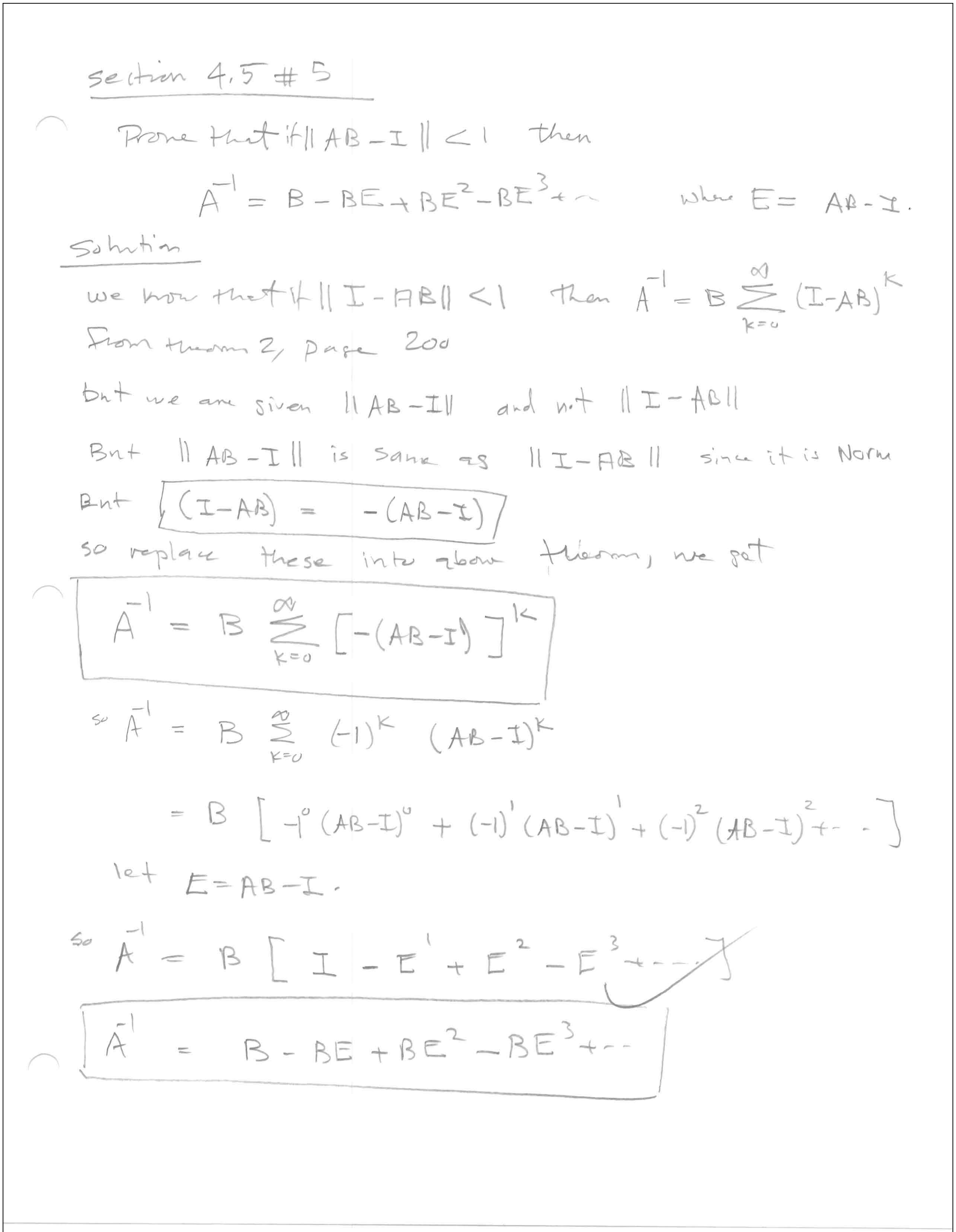
so $\|BA^{-1} - AA^{-1}\| < 1$

so $\|BA^{-1} - I\| < 1$ —— (1)

But from theorm 2, page 200, theorm on invertible matrices, it says that if $\|I - XY\| < 1$ then X, Y are invertible.

So compare this to (1), it means that $\underset{\overset{\downarrow}{X}}{B}$ and $\underset{\overset{\downarrow}{Y}}{A^{-1}}$ are both invertible.

$\Longrightarrow \boxed{B \text{ is invertible}}$



section 4.5 # 5

Prove that if $||AB - I|| < 1$ then

$$A^{-1} = B - BE + BE^2 - BE^3 + \cdots \quad \text{where } E = AB - I.$$

Solution

we know that if $||I - AB|| < 1$ then $A^{-1} = B \sum_{k=0}^{\infty} (I-AB)^k$

From theorm 2, page 200

But we are given $||AB - I||$ and not $||I - AB||$

But $||AB - I||$ is same as $||I - AB||$ since it is Norm

But $\boxed{(I-AB) = -(AB-I)}$

so replace these into above theorm, we get

$$\boxed{A^{-1} = B \sum_{k=0}^{\infty} \left[-(AB-I)\right]^k}$$

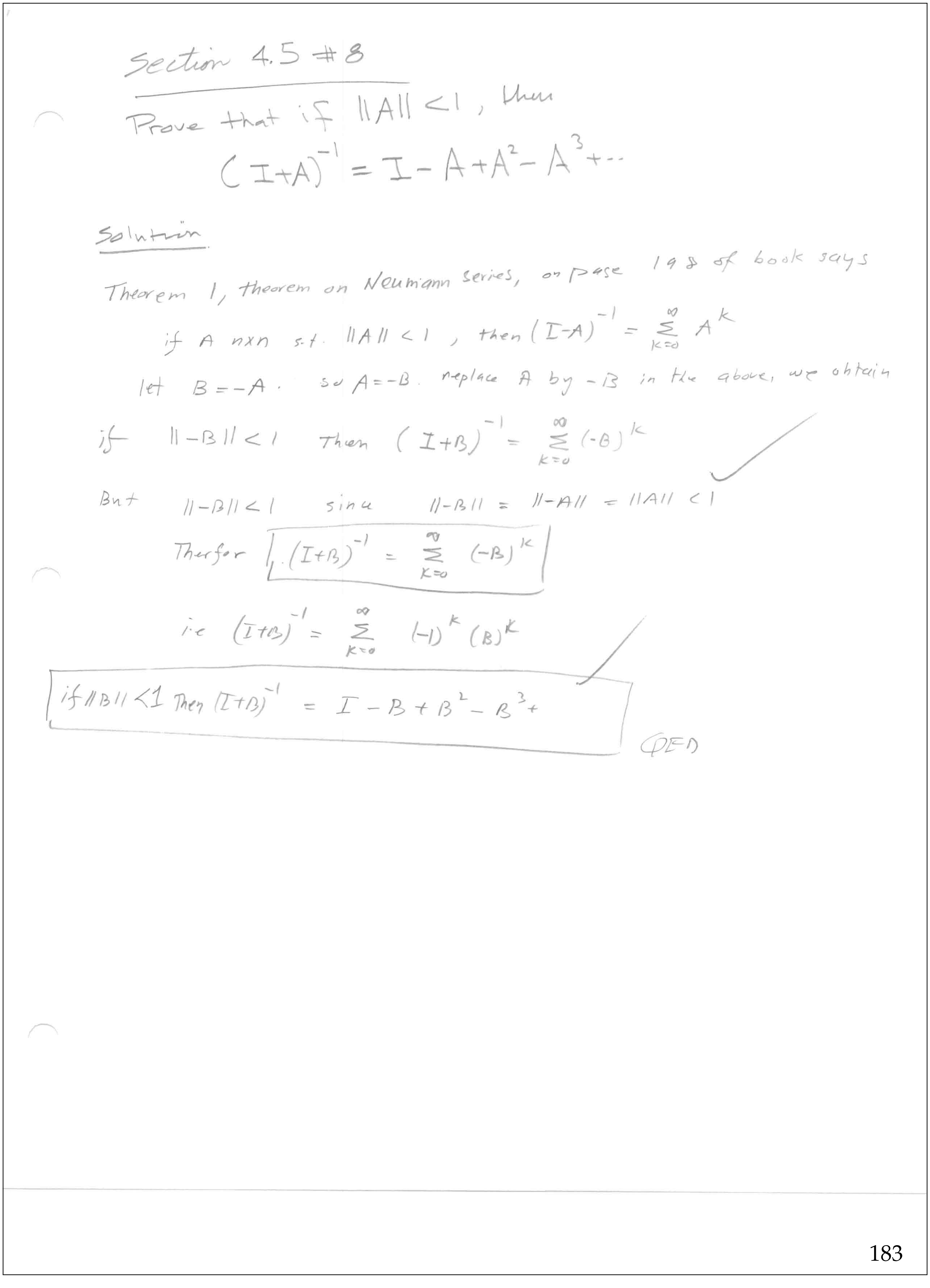
so $A^{-1} = B \sum_{k=0}^{\infty} (-1)^k \; (AB-I)^k$

$$= B \left[-1^0 (AB-I)^0 + (-1)^1 (AB-I)^1 + (-1)^2 (AB-I)^2 + \cdots \right]$$

let $E = AB - I$.

so $A^{-1} = B \left[I - E^1 + E^2 - E^3 + \cdots \right]$ ✓

$$\boxed{A^{-1} = B - BE + BE^2 - BE^3 + \cdots}$$



Section 4.5 #8

Prove that if $\|A\| < 1$, then

$$(I+A)^{-1} = I - A + A^2 - A^3 + \cdots$$

Solution

Theorem 1, theorem on Neumann series, on page 198 of book says

if A $n\times n$ s.t. $\|A\| < 1$, then $(I-A)^{-1} = \sum_{k=0}^{\infty} A^k$

let $B = -A$. so $A = -B$. replace A by $-B$ in the above, we obtain

if $\|-B\| < 1$ Then $(I+B)^{-1} = \sum_{k=0}^{\infty} (-B)^k$

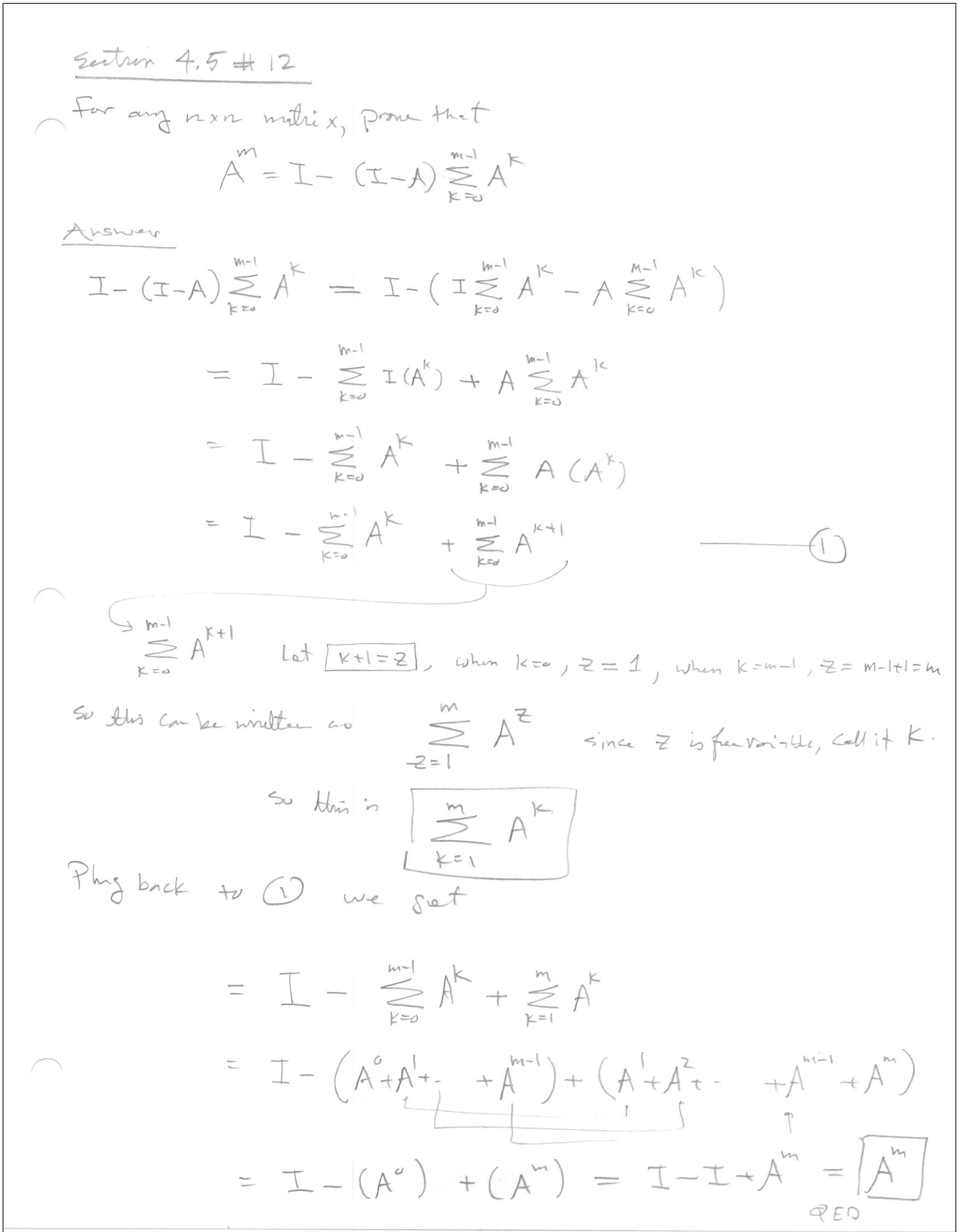
But $\|-B\| < 1$ since $\|-B\| = \|-A\| = \|A\| < 1$

Therfor $\boxed{(I+B)^{-1} = \sum_{k=0}^{\infty} (-B)^k}$

i.e $(I+B)^{-1} = \sum_{k=0}^{\infty} (-1)^k (B)^k$

$\boxed{\text{if } \|B\| < 1 \text{ Then } (I+B)^{-1} = I - B + B^2 - B^3 +}$

QED



Section 4.5 #12

For any n x n matrix, prove that

$$A^m = I - (I-A)\sum_{k=0}^{m-1} A^k$$

Answer

$$I-(I-A)\sum_{k=0}^{m-1}A^k = I-\left(I\sum_{k=0}^{m-1}A^k - A\sum_{k=0}^{m-1}A^k\right)$$

$$= I - \sum_{k=0}^{m-1} I(A^k) + A\sum_{k=0}^{m-1}A^k$$

$$= I - \sum_{k=0}^{m-1} A^k + \sum_{k=0}^{m-1} A(A^k)$$

$$= I - \sum_{k=0}^{m-1} A^k + \underbrace{\sum_{k=0}^{m-1} A^{k+1}} \text{ ——— } (1)$$

$\sum_{k=0}^{m-1} A^{k+1}$ Let $\boxed{k+1=z}$, when $k=0$, $z=1$, when $k=m-1$, $z=m-1+1=m$

So this can be written as $\sum_{z=1}^{m} A^z$ since z is free variable, call it k.

So this is $\boxed{\sum_{k=1}^{m} A^k}$

Plug back to (1) we get

$$= I - \sum_{k=0}^{m-1} A^k + \sum_{k=1}^{m} A^k$$

$$= I - \left(A^0 + A^1 + \cdots + A^{m-1}\right) + \left(A^1 + A^2 + \cdots + A^{m-1} + A^m\right)$$

$$= I - (A^0) + (A^m) = I - I + A^m = \boxed{A^m}$$

QED

section 4.5 #22

let $B_K = \sum_{j=0}^{K} A^j$. show that $[B_K]$ can be computed recursively by the formula $B_0 = I$, $B_{K+1} = I + AB_K$

solution

For $K=0$ $B_0 = A^0 = I$ so $\boxed{B_0 = I}$

For $K=1$

$$B_1 = A^0 + A$$
$$B_1 = I + A$$
$$= I + AI$$

but since $I = B_0$

$$\boxed{B_1 = I + AB_0}$$

For $K=2$

$$B_2 = A^0 + A^1 + A^2$$
$$= I + A(I + A)$$

but $I + A = B_1$ From above

so

$$\boxed{B_2 = I + AB_1}$$

For $K=3$

$$B_3 = A^0 + A^1 + A^2 + A^3$$
$$= I + A(I + A + A^2)$$
$$= I + A(I + A(I + A))$$

but $I + A = B_1$ From above

so

$$= I + A(I + AB_1)$$

but $I + AB_1 = B_2$ From above

$$\boxed{B_3 = I + AB_2}$$

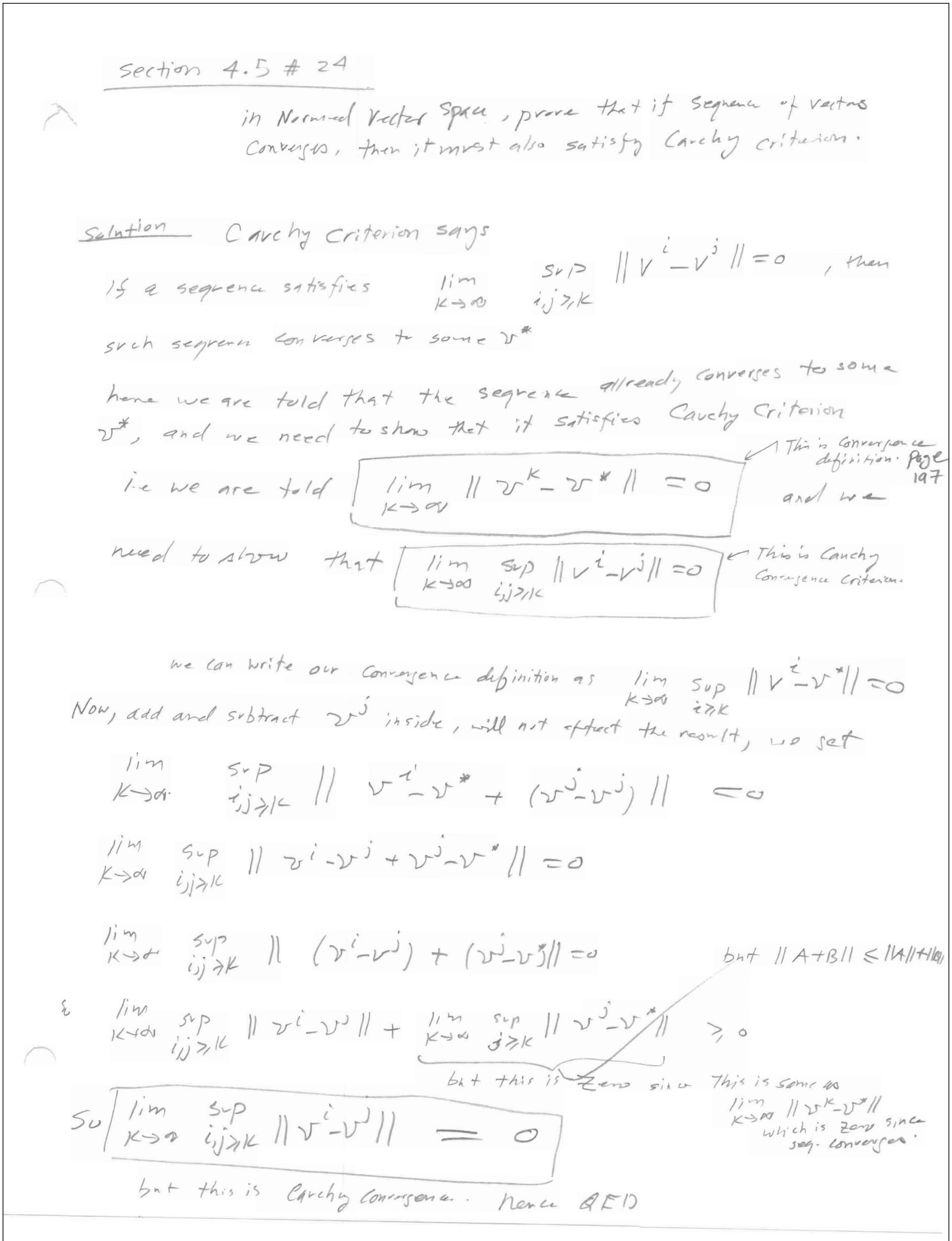
Therfor we see that the general recursive eqnation builds up as follows

$$B_0 = I$$
$$B_1 = I + AB_0$$
$$B_2 = I + AB_1$$
$$B_2 = I + AB_2$$
$$\vdots$$
$$\boxed{B_{K+1} = I + AB_K}$$

QED

This is proof by Construction. to show general Pattern.

Can also use proof by induction.



section 4.5 # 24

in Normed Vector Space, prove that if Sequence of Vectors Converges, then it must also satisfy Cauchy criterion.

Solution Cauchy Criterion says

If a sequence satisfies $\lim_{k\to\infty} \sup_{i,j\geq k} \|v^i - v^j\| = 0$, then such sequence converges to some v^*

here we are told that the sequence already converges to some v^*, and we need to show that it satisfies Cauchy Criterion

i.e we are told $\boxed{\lim_{k\to\infty} \|v^k - v^*\| = 0}$ This is Convergence definition. Page 197 and we need to show that $\boxed{\lim_{k\to\infty} \sup_{i,j\geq k} \|v^i - v^j\| = 0}$ This is Cauchy Convergence Criterion.

we can write our Convergence definition as $\lim_{k\to\infty} \sup_{i\geq k} \|v^i - v^*\| = 0$

Now, add and subtract v^j inside, will not effect the result, so get

$$\lim_{k\to\infty} \sup_{i,j\geq k} \| v^i - v^* + (v^j - v^j) \| = 0$$

$$\lim_{k\to\infty} \sup_{i,j\geq k} \| v^i - v^j + v^j - v^* \| = 0$$

$$\lim_{k\to\infty} \sup_{i,j\geq k} \| (v^i - v^j) + (v^j - v^*) \| = 0$$

but $\|A+B\| \leq \|A\| + \|B\|$

$$\lim_{k\to\infty} \sup_{i,j\geq k} \| v^i - v^j \| + \underbrace{\lim_{k\to\infty} \sup_{j\geq k} \| v^j - v^* \|}_{\text{but this is zero since}} \geq 0$$

This is same as $\lim_{k\to\infty} \|v^k - v^*\|$ which is zero since seq. converges.

So $\boxed{\lim_{k\to\infty} \sup_{i,j\geq k} \|v^i - v^j\| = 0}$

but this is Cauchy Convergence. hence QED

4.9 HW 8

Local contents

4.9.1 Analytic problems

4.9.1.1 section 4.6, problem 2

question: Prove that if A has this property (unit row diagonal dominant)

$$a_{ii} = 1 > \sum_{\substack{j=1\\j\neq i}}^{n} \left|a_{ij}\right| \qquad (1 \leq i \leq n)$$

then Richardson iteration is successful.

Solution:

Since the iterative formula is

$$\begin{aligned} x_{k+1} &= x_k + Q^{-1}\left(b - Ax_k\right) \\ &= x_k + Q^{-1}b - Q^{-1}Ax_k \\ &= (I - Q^{-1}A)x_k + Q^{-1}b \end{aligned}$$

This converges, by theorem (1) on page 210 when $\|I - Q^{-1}A\| < 1$

In Richardson method, $Q = I$, hence Richardson converges if $\|I - A\| < 1$

But $\|I - A\| = \| \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ 0 & \cdots & \ddots & \vdots \\ 0 & \cdots & \cdots & 1 \end{bmatrix} \begin{bmatrix} 1 & a_{12} & \cdots & a_{1n} \\ a_{21} & 1 & \cdots & a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 1 \end{bmatrix} \| = \| \begin{bmatrix} 0 & a_{12} & \cdots & a_{1n} \\ a_{21} & 0 & \cdots & a_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & 0 \end{bmatrix} \|$

But since row unit diagonal dominant, then the sum of each row elements remaining (after a_{ii} was annihilated) is a sum which is less than 1. Hence each row about will sum to some value which is less than 1. Hence the infinity norm of the above matrix, which is the maximum row sum, is less than 1. Hence

$$\|I - A\| < 1$$

Hence Richardson will converge. Each iteration will move closer to the solution x^*

4.9.1.2 problem 14

Problem: Prove that the eigenvalues of a Hermitian matrix are real.

Answer: A is Hermitian if $\overline{\left(A^T\right)} = A$, where the bar above indicates taking the complex conjugate. Hence the matrix is transposed and then each element will be complex conjugated.

Now, an eigenvalue of a matrix is defined such as

$$Ax = \lambda x$$

pre mutliply both sides by $\overline{\left(x^T\right)}$

$$\begin{aligned}\overline{\left(x^T\right)}Ax &= \overline{\left(x^T\right)}\lambda x\\ \overline{\left(x^T\right)}Ax &= \lambda\overline{\left(x^T\right)}x\end{aligned}$$

But since A is Herminitian, then $\overline{\left(A^T\right)} = A$, hence the above becomes

$$\begin{aligned}\overline{\left(x^T\right)\left(A^T\right)}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{\left(x^TA^T\right)}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{(Ax)^T}x &= \lambda\overline{\left(x^T\right)}x\end{aligned}$$

But $Ax = \lambda x$, hence the above becomes

$$\begin{aligned}\overline{(\lambda x)^T}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{\left(x^T\lambda\right)}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{\left(x^T\right)}\overline{\lambda}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{\lambda}\overline{\left(x^T\right)}x &= \lambda\overline{\left(x^T\right)}x\\ \overline{\lambda} &= \lambda\end{aligned}$$

Hence since complex conjugate of eigenvalue is the same as the eigenvalue, therefor λ is real.

4.9.1.3 problem 16

Problem: Prove that if A is nonsingular, then $A\overline{A^T}$ is positive definite.

Answer: A is nonsingular, meaning its left and right inverses exist and are the same. To show that a matrix is positive definite, need to show that $x^T Ax > 0$ for all $x \neq 0$.

Let $\overline{A^T} = B$, then let $n = x^T A\overline{A^T}x$, we need to show that $n > 0$

$$n = x^T (AB) x$$

Since A^{-1} exist, then multiply both sides by A^{-1} and B^{-1}

$$\begin{aligned} A^{-1}B^{-1}n &= A^{-1}B^{-1}x^T (AB) x \\ &= x^T x \end{aligned}$$

But $x^T x = \|x\|^2 > 0$ unless $x = 0$, hence above becomes

$$\begin{aligned} A^{-1}B^{-1}n &> 0 \\ A^{-1}\left(\overline{A^T}\right)^{-1} n &> 0 \\ \left(\overline{A^T}A\right)^{-1} n &> 0 \end{aligned}$$

Multiply both sides by $\overline{A^T}A$ leads to $n > 0$, hence $A\overline{A^T}$ is positive definite.

4.9.1.4 problem 17

Problem: Prove that if A is positive definite, then its eigenvalues are positive.

Answer: A is positive definite implies $x^T Ax > 0$ for all $x \neq 0$. i.e. $\langle x, Ax\rangle > 0$.

But $Ax = \lambda x$, hence

$$\langle x, \lambda x\rangle > 0$$

or

$$\lambda \langle x, x\rangle > 0$$

But $\langle x, x\rangle = \|x\|^2 > 0$ unless $x = 0$, therefor

$$\lambda > 0$$

4.9.1.5 section 4.7, problem 1

question: Prove that if A is symmetric, then the gradient of the function $q(x) = \langle x, Ax\rangle - 2\langle x, b\rangle$ at x is $2(Ax - b)$. Recall that the gradient of a function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is the vector whose components are $\frac{\partial g}{\partial x_i}$ for $i = 1, 2, \cdots n$

answer:

$\langle a, b\rangle$ is $a^T b$, hence using this definition, we can expend the RHS above and see that it will give the result required.

$$
\begin{aligned}
\langle x, Ax\rangle &= x^T(Ax)\\
&= [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n}\\ a_{21} & a_{22} & \cdots & a_{2n}\\ \vdots & \vdots & \ddots & \vdots\\ a_{n1} & a_{n2} & \cdots & a_{nn}\end{bmatrix}\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix}\\
&= [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n\\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n\\ \vdots\\ a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n\end{bmatrix}\\
&= x_1(a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + x_2(a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n)\\
&+ x_3(a_{31}x_1 + a_{32}x_2 + \cdots + a_{3n}x_n) + \cdots + x_n(a_{n1}x_1 + a_{n2}x_2 + \cdots + a_{nn}x_n)\\
&= \left(a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n\right) + \left(a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n\right)\\
&+ \left(a_{31}x_1x_3 + a_{32}x_2x_3 + a_{33}x_3^2 \cdots + a_{3n}x_nx_3\right) + \cdots + \left(a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2\right)
\end{aligned}
$$

And

$$
\begin{aligned}
\langle x, b\rangle &= x^T b\\
&= [x_1\ x_2\ \cdots\ x_n]\begin{bmatrix} b_1\\ b_2\\ \vdots\\ b_n\end{bmatrix}\\
&= x_1b_1 + x_2b_2 + \ \cdots + x_nb_n
\end{aligned}
$$

Hence Putting the above together, we obtain

$$
\begin{aligned}
q(x) &= \langle x, Ax\rangle - 2\langle x, b\rangle \\
&= \left(a_{11}x_1^2 + a_{12}x_1x_2 + \cdots + a_{1n}x_1x_n\right) + \left(a_{21}x_2x_1 + a_{22}x_2^2 + \cdots + a_{2n}x_2x_n\right) \\
&\quad + \left(a_{31}x_1x_3 + a_{32}x_2x_3 + a_{33}x_3^2 + \cdots + a_{3n}x_nx_3\right) + \cdots + \left(a_{n1}x_nx_1 + a_{n2}x_nx_2 + \cdots + a_{nn}x_n^2\right) \\
&\quad - 2\left(x_1b_1 + x_2b_2 + \ \cdots + x_nb_n\right)
\end{aligned}
$$

Now taking the derivative of the above w.r.t $x_1, x_2, \cdots, x_n$ to generate the gradient vector, we obtain

$$
\begin{aligned}
\frac{\partial q(x)}{\partial x_1} &= (2a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n) + (a_{21}x_2) + (a_{31}x_3) + \cdots + (a_{n1}x_n) - 2(b_1) \\
\frac{\partial q(x)}{\partial x_2} &= (a_{12}x_1) + (a_{21}x_1 + 2a_{22}x_2 + \cdots + a_{2n}x_n) + (a_{32}x_3) + \cdots + (a_{n2}x_n) - 2(b_2) \\
&\vdots \\
\frac{\partial q(x)}{\partial x_n} &= (a_{1n}x_1) + (a_{2n}x_2) + (a_{3n}x_3) + \cdots + (a_{n1}x_1 + a_{n2}x_2 + \cdots + 2a_{nn}x_n) - 2(b_n)
\end{aligned}
$$

Hence

$$
\begin{aligned}
\frac{\partial q(x)}{\partial x_1} &= \left[2a_{11},\ a_{12,}\ \cdots, a_{1n}\right]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} + [0\ , a_{21}, a_{31},\ \cdots, a_{n1}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} - 2(b_1) \\
\frac{\partial q(x)}{\partial x_2} &= [a_{21}\ , 2a_{22}\ , \cdots, a_{2n}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} + [a_{12}\ , 0,\ \cdots, a_{n2}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} - 2(b_2) \\
&\vdots \\
\frac{\partial q(x)}{\partial x_n} &= [a_{n1}\ , a_{n2},\ \cdots, 2a_{nn}]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} + \left[a_{n1},\ a_{n2,}\ \cdots, 0\right]\begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix} - 2(b_n)
\end{aligned}
$$

Combine, we get

$$
\begin{aligned}
\frac{\partial q(x)}{\partial x_1} &= \left[2a_{11},\ 2a_{12,}\ \cdots,2a_{1n}\right]\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} - 2\left(b_1\right)\\
\frac{\partial q(x)}{\partial x_2} &= [2a_{21}\ ,2a_{22}\ ,\cdots,2a_{2n}]\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} - 2\left(b_2\right)\\
&\vdots\\
\frac{\partial q(x)}{\partial x_n} &= [2a_{n1}\ ,2a_{n2},\ \cdots,2a_{nn}]\begin{bmatrix} x_1\\ x_2\\ \vdots\\ x_n\end{bmatrix} - 2\left(b_n\right)
\end{aligned}
$$

Hence, in Vector/Matrix notation we obtain

$$
\begin{aligned}
\frac{\partial q(x)}{\partial \vec{x}} &= 2A\vec{x} - 2\vec{b}\\
&= 2\left(A\vec{x} - \vec{b}\right)
\end{aligned}
$$

Which is what we are asked to show.

(it would also have been possible to solve the above by expressing everything in the summation notations, i.e. writing $(Ax)_i = \sum_j^n A\left(i,j\right)x_j$, and apply differentiations directly on these summation expression as is without expanding them as I did above, it would have been probably shorter solution)

4.9.1.6 problem 2

Question: Prove that the minimum value of $q(x)$ is $-\left\langle b, A^{-1}b\right\rangle$

Solution:

$$q(x) = \langle x, Ax\rangle - 2\langle x, b\rangle$$

From problem (1) we found that $\frac{\partial q(x)}{\partial \vec{x}} = 2\left(A\vec{x} - \vec{b}\right)$, hence setting this to zero

$$\begin{aligned} A\vec{x} - \vec{b} &= 0 \\ \vec{x} &= A^{-1}b \end{aligned}$$

To check if this is a min,max, or saddle, we differentiate $\frac{\partial q(x)}{\partial \vec{x}}$ once more, and plug in this solution and check

$$\begin{aligned} \frac{\partial}{\partial \vec{x}}\left(\frac{\partial q(x)}{\partial \vec{x}}\right) &= 2\frac{\partial}{\partial \vec{x}}\left(A\vec{x} - \vec{b}\right) \\ &= 2A \end{aligned}$$

(A needs to be positive definite here, problem did not say this?) Hence this is a minimum.

Hence minimum value $q(x)$ is

$$q_{\min}(x) = \left\langle A^{-1}b, Ax\right\rangle - 2\left\langle A^{-1}b, b\right\rangle$$

But $Ax = b$, hence

$$\begin{aligned} q_{\min}(x) &= \left\langle A^{-1}b, b\right\rangle - 2\left\langle A^{-1}b, b\right\rangle \\ &= -\left\langle A^{-1}b, b\right\rangle \\ &= -\left\langle b, A^{-1}b\right\rangle \end{aligned}$$

4.9.1.7 problem 6

question: Prove that if $\hat{t} = \frac{\langle v,b-Ax\rangle}{\langle v,Av\rangle}$,and if $y = x + \hat{t}v$ then $\langle v, b - Ay\rangle = 0$

answer:

$$y = x + \frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle} v$$

Then

$$\begin{aligned}
\langle v, b - Ay\rangle &= \left\langle v, b - A\left(x + \frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle} v\right)\right\rangle \\
&= \left\langle v, b - Ax - A\frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle} v\right\rangle \\
&= \left\langle v, b - Ax - Av\frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle}\right\rangle \\
&= v^T\left(b - Ax - Av\frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle}\right) \\
&= v^T b - v^T Ax - v^T Av\frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle} \\
&= v^T b - v^T Ax - \langle v, Av\rangle \frac{\langle v, b - Ax\rangle}{\langle v, Av\rangle} \\
&= v^T b - v^T Ax - \langle v, b - Ax\rangle \\
&= v^T (b - Ax) - \langle v, b - Ax\rangle \\
&= \langle v, b - Ax\rangle - \langle v, b - Ax\rangle \\
&= 0
\end{aligned}$$

4.9.2 Computer problems

4.9.2.1 Jacobi, Gauss-Seidel, SOR

Jacobi, Gauss-Seidel, and SOR iterative solvers were implemented (in Matlab) and compared for rate of convergence. The implementation was tested on the following system

$$A = \begin{bmatrix} -4 & 2 & 1 & 0 & 0 \\ 1 & -4 & 1 & 1 & 0 \\ 2 & 1 & -4 & 1 & 2 \\ 0 & 1 & 1 & -4 & 1 \\ 0 & 0 & 1 & 2 & 04 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix} = \begin{bmatrix} -4 \\ 11 \\ -16 \\ 11 \\ -4 \end{bmatrix}$$

and the result was compared to Matlab $A\backslash b'$ result, which is

```
A\b'
ans =
   1.00000000000000
  -2.00000000000000
   4.00000000000000
  -2.00000000000000
   1.00000000000000
>>
```

These iterative solvers solve the linear system $Ax = b$ by iteratively approaching a solution. In all of these methods, the general iteration formula is

$$x_{k+1} = x_k + Q^{-1}(I - Ax_k)$$

Each method uses a different Q matrix. For Jacobi, $Q = diag(A)$, and for Gauss-Seidel, $Q = L(A)$, which is the lower triangular part of A. For SOR, $Q = \frac{1}{\omega}\left(diag(A) + \omega L^0(A)\right)$ where $L^0(A)$ is the strictly lower triangular part of A, and ω is an input parameter generally $0 < \omega < 2$.

For the Jacobi and Gauss-Seidel methods, we are guaranteed to converge to a solution if A is diagonally dominant. For SOR the condition of convergence is $\rho(G) < 1$, where $G = \left(I - Q^{-1}A\right)$, and $\rho(G)$ is the spectral radius of G.

4.9.2.2 Results

The following table shows the result of running Jacobian and Gauss-Seidel on the same Matrix. The columns are the relative error between each successive iteration. Defined as $\frac{|x_{k+1}-x_k|}{|x_{k+1}|}$. The first column is the result from running the Jacobian method, and the second is the result from the Gauss-Seidel method.

Iteration	Jacobi	Gauss-Seidel
1	0.94197873843414	0.93435567469174
2	0.22011159808466	0.13203098317070
3	0.04613055928361	0.03056375635161
4	0.02582530967034	0.02878251571226
5	0.02166103846735	0.02294426703511
6	0.01508738881859	0.01935275191913
7	0.01447950687767	0.01618762392092
8	0.01246529285252	0.01353611057259
9	0.01166876946105	0.01130586303057
10	0.01055706870353	0.00943545181644
11	0.00971262454554	0.00786920716886
12	0.00886458421973	0.00655940732143
13	0.00811604581680	0.00546521547469
14	0.00741643318343	0.00455191661070
15	0.00677994422374	0.00379012917894
16	0.00619437598238	0.00315507291046
17	0.00565882948442	0.00262590555197
18	0.00516810821828	0.00218513512307
19	0.00471915312779	0.00181810678260
20	0.00430837834153	0.00151255968550

We see from the above table that Gauss-Seidel method has faster convergence than Jacobi method. The following is a plot of the above table.

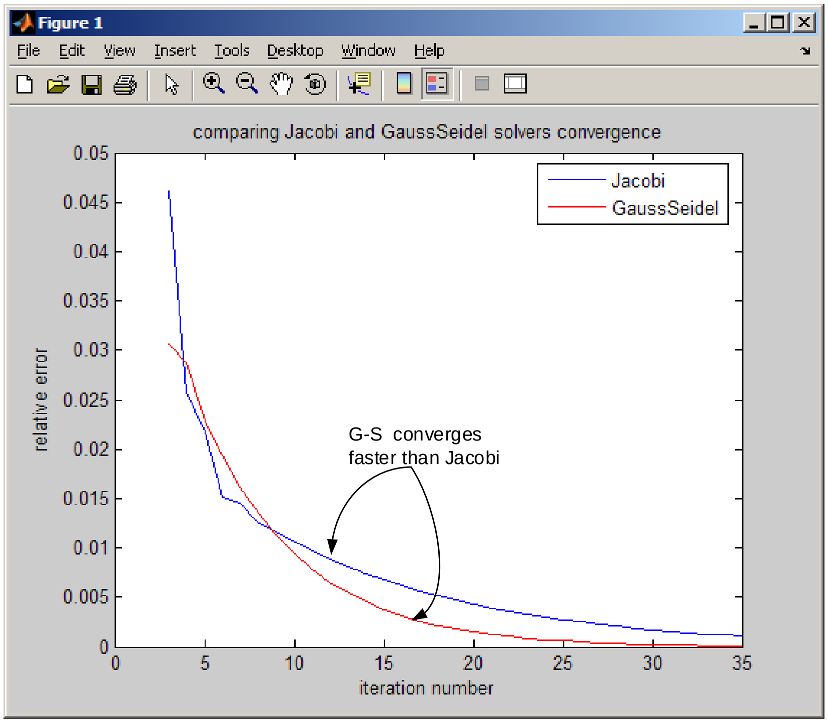


Figure 4.3: Plot

For the SOR method, it depends on the value of ω used. First we look at SOR on its own, comparing its performance as ω changes. Next, we pick 2 values for ω and compare SOR using these values with the Jacobian and the Gauss-Seidel method.

This table below shows the values of the relative error for difference ω. Only the first 20 iterations are shown. The following ω values are used: .25, .5, .75, 1, 1.25, 1.5, 1.75. This table also shows the number of iterations needed to achieve the same error tolerance specified by the user. Smaller number of iterations needed to reach this error limit indicates that ω selected was better than otherwise.

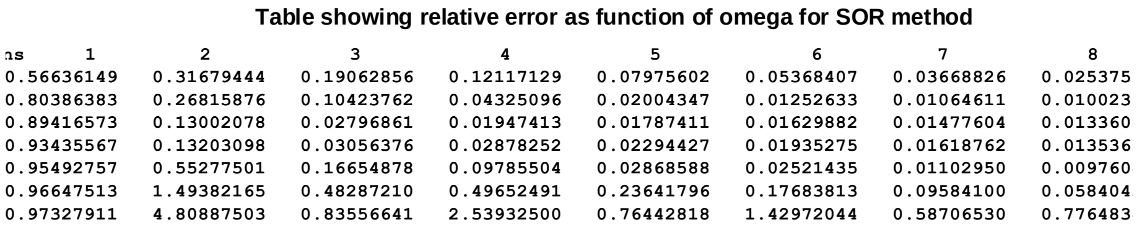


Table showing relative error as function of omega for SOR method

ns 1	2	3	4	5	6	7	8
0.56636149	0.31679444	0.19062856	0.12117129	0.07975602	0.05368407	0.03668826	0.025375
0.80386383	0.26815876	0.10423762	0.04325096	0.02004347	0.01252633	0.01064611	0.010023
0.89416573	0.13002078	0.02796861	0.01947413	0.01787411	0.01629882	0.01477604	0.013360
0.93435567	0.13203098	0.03056376	0.02878252	0.02294427	0.01935275	0.01618762	0.013536
0.95492757	0.55277501	0.16654878	0.09785504	0.02868588	0.02521435	0.01102950	0.009760
0.96647513	1.49382165	0.48287210	0.49652491	0.23641796	0.17683813	0.09584100	0.058404
0.97327911	4.80887503	0.83556641	2.53932500	0.76442818	1.42972044	0.58706530	0.776483

Figure 4.4: Table

From the above table, and for the specific data used in this exercise, we observe that $\omega = 1.25$ achieved the best convergence

This is a plot of the above table. (using zoom to make the important part more visible).

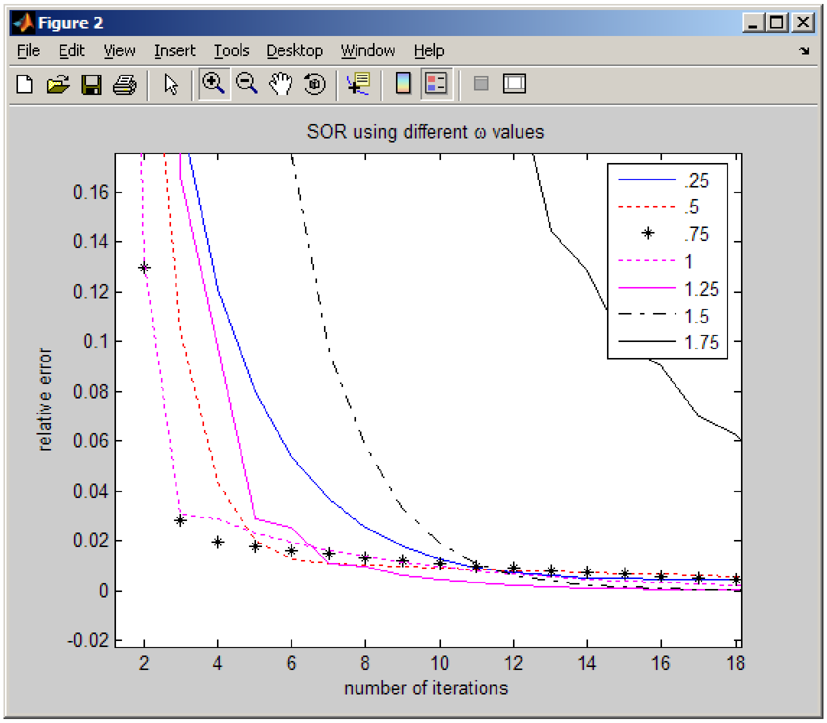


Figure 4.5: Zoom plot

Now we show the difference between Jacobi, Gauss-Seidel and SOR (using $\omega = 1.25$) as this ω gave the best convergence using SOR. The following is a plot showing that SOR with $\omega = 1.25$ achieved best convergence.

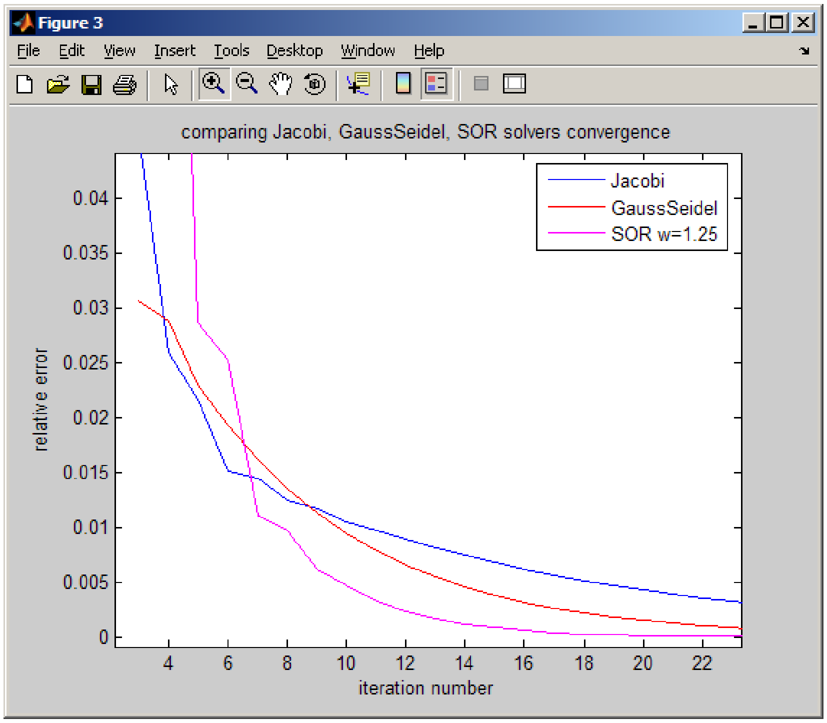


Figure 4.6: SOR plot

4.9.2.3 Example Matlab output

The following is an output from one run generated by a Matlab script written to test these implementations. 2 test cases used. A,b input as given in the HW assignment sheet given in the class, and a second A,b input shown in the textbook (SPD matrix, but not diagonally dominant) shown on page 245. This is the output.

```
***** TEST 1 *********
A =
    -4     2     1     0     0
     1    -4     1     1     0
     2     1    -4     1     2
     0     1     1    -4     1
     0     0     1     2    -4


b =
    -4    11   -16    11    -4
======>Matlab linear solver solution, using A\b
   1.00000000000000
  -2.00000000000000
   4.00000000000000
  -2.00000000000000
   1.00000000000000
```

```
Solution from Jacobi
   1.00037323222607
  -1.99962676777393
   4.00061424742539
  -1.99962676777393
   1.00037323222607

Solution from Gauss-Seidel
   1.00019802638995
  -1.99981450377552
   4.00028764196082
  -1.99983534139755
   1.00015423979143

Solution from SOR, w=0.250000
   1.00355389692825
  -1.99647784171454
   4.00574940211158
  -1.99653496988616
   1.00343408511030

Solution from SOR, w=0.500000
   1.00064715564763
  -1.99936629645376
   4.00102314817150
  -1.99939010276819
   1.00059721960251

Solution from SOR, w=0.750000
   1.00036324343971
  -1.99965036525822
   4.00055574249388
  -1.99967387439430
   1.00031390750518

Solution from SOR, w=1.000000
   1.00019802638995
  -1.99981450377552
   4.00028764196082
  -1.99983534139755
   1.00015423979143

Solution from SOR, w=1.250000
   1.00009213101077
  -1.99991820839097
   4.00012079232435
  -1.99993416539569
```

```
   1.00005844632680

Solution from SOR, w=1.500000
   0.99998926445485
  -1.99999117155337
   3.99998500044858
  -1.99999354920373
   0.99999484763949

Solution from SOR, w=1.750000
   1.00001519677595
  -2.00001369840580
   4.00002560215918
  -2.00001192126720
   1.00001112407411

***** TEST 2 *********
A =
    10     1     2     3     4
     1     9    -1     2    -3
     2    -1     7     3    -5
     3     2     3    12    -1
     4    -3    -5    -1    15

b =
    12   -27    14   -17    12

======>Matlab linear solver solution, using A\b
   1.00000000000000
  -2.00000000000000
   3.00000000000000
  -2.00000000000000
   1.00000000000000

Jacobi solution
   1.00018168868849
  -2.00016761404521
   2.99969203212914
  -1.99995200794879
   0.99978228017035

Gauss Seidel solution
   1.00009837024472
  -2.00008075631740
   2.99986024703377
  -1.99998691907114
   0.99991190441111
```

```
SOR solution, w=1.25
   1.00002736848705
  -2.00001960859025
   2.99996840603470
  -1.99999920522245
   0.99998285315310
```

4.9.2.3.1 Conclusion

SOR with $\omega = 1.25$ achieved the best convergence. However, one needs to find determine which ω can do the best job, and this also will depend on the input. For different input different ω might give better result. Hence to use SOR one must first do a number of trials to determine the best value to use. Between the Jacobian and the Gauss-Seidel methods, the Gauss-Seidel method showed better convergence.

4.9.2.4 Long operation count

In these operations long count, since these algorithms are iterative, hence the operation count will be based on one iteration. One would then need to multiply this count by the number of iterations need to converge as per the specification that the user supplies. Gauss-Seidel will require less iterations to converge to the same solution as compared to Jacobi method. With the SOR method, it will depend on the ω selected.

4.9.2.4.1 Jacobi

Long operations count for one iteration of the Jacobi method.

```
while keepLookingForSolution
   k=k+1;
   xnew=xold+Qinv*(b-A*xold);- - - - - (1)

   currentError=norm(xnew-xold);- - - - - (2)
   relError(k)=currentError/norm(xnew);- - - - - (3)

   if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
       keepLookingForSolution=FALSE;
   else
     xold=xnew;
   end

end
```

where

$$Qinv = eye(nRow)/diag(diag(A));$$

For line(1): For multiplying an $n \times n$ matrix by $n \times 1$ vector, n^2 ops are required. Hence for $A * xold$ we require n^2 ops. The result of $(b - A * xold)$ is an n long vector. Next we have $Qinv$ multiplied by this n long vector. This will require only n ops since $Qinv$ is all zeros except at the diagonal. Hence to calculate xnew we require $n + n^2$ ops.

For line(2,3): It involves 2 norm operations and one division. Hence we need to find the count for the norm operation. Assuming norm 2 is being used which is defined as $\sqrt{\sum_{i=1}^{n} x_i^2}$ hence this requires n multiplications and one long operation added for taking square root (ok to do?). Hence the total ops for finding relative error is $2n + 2 + 1 = 2n + 3$

The next operation is the check for the result limit:

$$norm(b - A * xnew) <= resLimit$$

This involves n^2 operations for the matrix by vector multiplication, and $n + 1$ operations for the norm. Hence $n^2 + n + 1$

Adding all the above we obtain: $n + n^2 + 2n + 3 + n^2 + n + 1 = 2n^2 + 4n + 4$ hence this is $O\left(2n^2\right)$

The above is the cost per one iteration. Hence if M is close to n, then it is worst than G.E. But if the number of iteration is much less than N, then this method will be faster than non-iterative methods based on Gaussian elimination which required $O\left(n^3\right)$.

4.9.2.4.2 Gauss-Seidel Long operations count for one iteration of the Gauss-Seidel. For the statement

```
 while keepLookingForSolution
    k=k+1;

    xnew=xold;
    for i=1:nRow
        xnew(i)=xnew(i)+ (b(i)-A(i,:)*xnew)/A(i,i);- - - - -(1)
    end

    currentError=norm(xnew-xold); - - - - -(2)
    relError(k)=currentError/norm(xnew); - -- - - -(3)

    if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
        keepLookingForSolution=FALSE;
    else
        xold=xnew;
    end
end
```

For line (1): For each i, there is n for the operation A(i,:)*xnew, and one operation for the division. Hence the total cost is n^2 since there are n rows.

For line(2+3): The cost is the same as with the Jacobi method, which was found to be $2n + 3$

The next operation is the check for the result limit:

$$norm(b - A * xnew) <= resLimit$$

This involves n^2 operations for the matrix by vector multiplication, and $n+1$ operations for the norm.

Hence the total operation is $n^2 + 2n + 3 + n + 1 = \boxed{n^2 + 3n + 4}$ hence this is $O\left(n^2\right)$ per one iteration.

4.9.2.4.3 SOR

This is the same as Gauss-Seidel, except there is an extra multiplication by ω per one row per one iteration. Hence we need to add n to the count found in Gauss-Seidel. Hence the count for SOR is $\boxed{n^2 + 4n + 4}$ or $O\left(n^2\right)$

4.9.2.5 Steepest descent iterative solver algorithm

The following is the output from testing the Steepest descent iterative algorithm.

```
======>Matlab linear solver solution, using A\b
   1.00000000000000
  -2.00000000000000
   3.00000000000000
  -2.00000000000000
   1.00000000000000

Steepest descent solution
   1.00013109477547
  -2.00012625234312
   2.99976034511258
  -1.99996859964602
   0.99987563548937

>>
```

4.9.2.6 Steepest Descent operation count

```
while keepLookingForSolution
    k=k+1;

    v=b-A*xold;    - - - - - - (1)
    t=dot(v,v)/dot(v,A*v); - - - - -(2)
    xnew=xold+t*v; - - - - -(3)

    currentError=norm(xnew-xold);   - - - - -(4)
    relError(k)=currentError/norm(xnew); - - - -(5)

    if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
```

```
        keepLookingForSolution=FALSE;
    else
        xold=xnew;
    end
end
```

For line (1): n^2

For line (2): For numerator: dot operation is n. For denominator: for A*v need n^2, and then need n added to the dot operation, hence need $n + n^2 + n$ operations. Add 1 to the division, hence need $n^2 + 2n + 1$ for line (2).

For line(3): n multiplications.

For line(4): $n + 1$ for the norm.

For line(5): $n + 2$ operations.

And finally for norm(b-A*xnew) in the final check, requires $n^2 + n + 1$

Hence total is $n^2 + n^2 + 2n + 1 + n + n + 1 + n + 2 + n^2 + n + 1 = 3n^2 + 6n + 5$

4.9.3 Source code

4.9.3.1 nma_driverTestIterativeSolvers.m

```
%
%This script is the driver to test and gather data for plotting
%for computer assignment 3/19/07 for Math 501, CSUF
%
%Nasser Abbasi 032607
%
%file name: nma_driverTestIterativeSolvers.m

close all;
clear all;

DISP_FOR_TABLE=0;  %turn to 1 to get output for table display

A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
b=[-4 11 -16 11 -4];
maxIter=200;
errorLimit=0.0001;
resLimit=0.00001;
oTable=zeros(maxIter,2);

%nma_getSpectraRadiusOfMatrix(
%find spectral radius for (I-Q^-1 A)

Q=diag(diag(A));
```

```
r=max(abs(eig(  eye(size(A,1)) - inv(Q)*A  ) ));
fprintf('Jacboi: spectral radius is %f\n',r);
if r>1
   fprintf('WARNING, spectral radius of (I-Q^-1 A)  should be less than 1 for convergence\n')
end
[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',  ...
    maxIter,errorLimit,resLimit);
figure;
plot(3:35,relError(3:35)); % plot(relError(1:k));
oTable(1:k,1)=relError(1:k);
fprintf('Solution from Jacobi\n');
format long;
disp(x)


Q=tril(A);
r=max(abs(eig(  eye(size(A,1)) - inv(Q)*A  ) ));

fprintf('Jacboi: spectral radius is %f\n',r);
if r>1
   fprintf('WARNING, spectral radius of (I-Q^-1 A)  should be less than 1 for convergence\n')
end

[x,k,relError]=nma_GaussSeidelIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
hold on;
plot(3:35,relError(3:35),'r'); % plot(relError(1:k));
legend('Jacobi','GaussSeidel');
title('comparing Jacobi and GaussSeidel solvers convergence');
xlabel('iteration number');
ylabel('relative error');

fprintf('Solution from Gauss-Seidel\n');
format long;
disp(x)


oTable(1:k,2)=relError(1:k);
fprintf('J\tG-S\n');
for i=1:35
    fprintf('%d\t%16.15f\t%16.15f\n',i,oTable(i,1),oTable(i,2));
end

%do it again for inclusion into Latex
if DISP_FOR_TABLE
    fprintf('Jacob\n');
    format long;
    for i=1:20
        disp(oTable(i,1))
    end

    %do it again for inclusion into Latex
    fprintf('G-S\n');
    for i=1:20
        disp(oTable(i,2))
    end
end

figure;
omegaValues=0.25*[1 2 3 4 5 6 7];
oTable=zeros(length(omegaValues),maxIter+1);
mycolor={'b','r:','k*','m:','m','k-.','k'};
for i=1:length(omegaValues)
    [x,k,relError]=nma_SORIterativeSolver(A,[1,1,1,1,1]',b',...
        maxIter,errorLimit,resLimit,omegaValues(i));
    plot(relError(1:k),mycolor{i});
    oTable(i,1)=omegaValues(i);
    oTable(i,2)=k;
    oTable(i,3:3+k-1)=relError(1:k);
    hold on;
    fprintf('Solution from SOR, w=%f\n',omegaValues(i));
    disp(x)

end
title('SOR using different \omega values');
xlabel('number of iterations');
ylabel('relative error');
legend('.25','.5','.75','1','1.25','1.5','1.75');

fprintf('omega\titeration\trelative errors\n');
for i=1:length(omegaValues)
    fprintf('%3.2f\t%d',oTable(i,1),oTable(i,2));
    if(oTable(i,2)>35)
        cutOff=35;
    else
        cutOff=oTable(i,2);
    end
    fprintf('\t\t');
    for j=1:cutOff
        fprintf('\t%16.15f',oTable(i,j+2));
        %fprintf('   %9.8f',oTable(i,j+2));
    end
    fprintf('\n');
end
```

```
%do it again for inclusion into Latex
if DISP_FOR_TABLE
    fprintf('SOR\n');
    for j=1:length(omegaValues)
        fprintf('SOR======>\n');
        for i=1:20
            disp(oTable(j,i+2))
        end
    end
end



%now compare the 3 methods using omega 1.25 only
[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
figure;
plot(3:35,relError(3:35)); % plot(relError(1:k));
[x,k,relError]=nma_GaussSeidelIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
hold on;
plot(3:35,relError(3:35),'r'); % plot(relError(1:k));

[x,k,relError]=nma_SORIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit,1.25);
plot(3:35,relError(3:35),'m'); % plot(relError(1:k));

legend('Jacobi','GaussSeidel','SOR w=1.25');
title('comparing Jacobi, GaussSeidel, SOR solvers convergence');
xlabel('iteration number');
ylabel('relative error');


%%%% Now run the test again. short version to paste into document.


A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
b=[-4 11 -16 11 -4];

fprintf('***** TEST 1 *********\n');
A
b
fprintf('======>Matlab linear solver solution, using A\\b  \n');
disp(A\b');

[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
fprintf('Jacobi solution\n');
disp(x);

[x,k,relError]=nma_GaussSeidelIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
fprintf('Gauss Seidel solution\n');
disp(x);

[x,k,relError]=nma_SORIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit,1.25);
fprintf('SOR solution, w=1.25\n');
disp(x);



%%%% Now run another test. Use an SPD matrix, which is shown on
%page 245 of textbook (Numerical Analysis, Kincaid.Cheney)
fprintf('***** TEST 2 *********\n');
A=[10 1 2 3 4;1 9 -1 2 -3;2 -1 7 3 -5;3 2 3 12 -1;4 -3 -5 -1 15];
b=[12 -27 14 -17 12];
A
b
fprintf('======>Matlab linear solver solution, using A\\b  \n');
disp(A\b');

[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
fprintf('Jacobi solution\n');
disp(x);

[x,k,relError]=nma_GaussSeidelIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit);
fprintf('Gauss Seidel solution\n');
disp(x);

[x,k,relError]=nma_SORIterativeSolver(A,[1,1,1,1,1]',b',...
    maxIter,errorLimit,resLimit,1.25);
fprintf('SOR solution, w=1.25\n');
disp(x);
```

4.9.3.2 nma_SORIterativeSolver.m

```
function [xnew,k,relError]=nma_SORIterativeSolver(A,x,b, ...
    maxIter,errorLimit,resLimit,omega)
%function [xnew,k,relError]=nma_GaussSeidelIterativeSolver(A,x,b,...
%                                       maxIter,errorLimit,resLimit,omega)
%
% Solve Ax=b using the SOR Iterative method
%
%INPUT:
% A: the A matrix
% x: Initial guess for solution
% b: right hand side
% maxIter:  max number of iterations allowed
% errorLimit: error tolerance. difference between successive x iteration
%             values. if such a difference is less than this error, stop.
% resLimit: if |b-A*x| is less than this limit, stop the iterative process.
% omega: SOR factor
%
%OUTPUT
% xnew: the solution found by iterative method.
% k: actual number of iterations used to obtain the above solution.
% relError: array that contains the relative error found at each iteration

%example call
% A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
% b=[-4 11 -16 11 -4]; maxIter=200; errorLimit=0.0001; resLimit=0.00001;
%[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',...
%                                          maxIter,errorLimit,resLimit)

%by Nasser Abbasi  3/26/07
%

% do some error checking on input....

if nargin ~=7
    error 'wrong number of arguments. 7 inputs are required';
end

if ~isnumeric(omega)
    error 'omega must be numeric';
end

TRUE=1; FALSE=0;

[res,msg]=nma_IterativeSolversIsValidInput(A,x,b,...
    maxIter,errorLimit,resLimit);
if ~res
```

```
    error(msg);
end

[nRow,nCol]=size(A);
xold=x(:);
b=b(:);
k=0;
relError=zeros(maxIter,1);

keepLookingForSolution=TRUE;

while keepLookingForSolution
    k=k+1;

    xnew=xold;
    for i=1:nRow
        xnew(i)=xnew(i)+ omega*(b(i)-A(i,:)*xnew)/A(i,i);
    end

    currentError=norm(xnew-xold);
    relError(k)=currentError/norm(xnew);

    if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
        keepLookingForSolution=FALSE;
    else
        xold=xnew;
    end
end

end
```

4.9.3.3 nma_JacobiIterativeSolver.m

```
function [xnew,k,relError]=nma_JacobiIterativeSolver(A,x,b, ...
                                            maxIter,errorLimit,resLimit)
%function [xnew,k]=nma_JacobiIterativeSolver(A,x,b,...
%                                           maxIter,errorLimit,resLimit)
%
% Solve Ax=b using the Jacobi Iterative method
%
%INPUT:
% A: the A matrix
% x: Initial guess for solution
% b: right hand side
% maxIter:  max number of iterations allowed
% errorLimit: error tolerance. difference between successive x iteration
%             values. if such a difference is less than this error, stop.
% resLimit: if |b-A*x| is less than this limit, stop the iterative process.
%
%OUTPUT
% xnew: the solution found by iterative method.
% k: actual number of iterations used to obtain the above solution.
% relError: array that contains the relative error found at each iteration
%
%example call
% A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
% b=[-4 11 -16 11 -4]; maxIter=200; errorLimit=0.0001; resLimit=0.00001;
%[x,k,relError]=nma_JacobiIterativeSolver(A,[1,1,1,1,1]',b',...
%                                          maxIter,errorLimit,resLimit)

%by Nasser Abbasi  3/26/07
%

if nargin ~=6
    error 'wrong number of arguments. 6 inputs are required';
end

TRUE=1; FALSE=0;

[res,msg]=nma_IterativeSolversIsValidInput(A,x,b,maxIter,errorLimit,resLimit);
if ~res
    error(msg);
end

[nRow,nCol]=size(A);
xold=x(:);
b=b(:);
k=0;
relError=zeros(maxIter,1);
Qinv=eye(nRow)/diag(diag(A));

keepLookingForSolution=TRUE;

while keepLookingForSolution
   k=k+1;
   xnew=xold+Qinv*(b-A*xold);

   currentError=norm(xnew-xold);
   relError(k)=currentError/norm(xnew);

   if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
       keepLookingForSolution=FALSE;
   else
     xold=xnew;
   end

end

end
```

4.9.3.4 nma_GaussSeidelIterativeSolver.m

```
function [xnew,k,relError]=nma_GaussSeidelIterativeSolver(A,x,b, ...
    maxIter,errorLimit,resLimit)
%function [xnew,k]=nma_GaussSeidelIterativeSolver(A,x,b,...
%                                             maxIter,errorLimit,resLimit)
%
% Solve Ax=b using the Gauss-Seidel Iterative method
%
%INPUT:
% A: the A matrix
% x: Initial guess for solution
% b: right hand side
% maxIter:  max number of iterations allowed
% errorLimit: error tolerance. difference between successive x iteration
%             values. if such a difference is less than this error, stop.
% resLimit: if |b-A*x| is less than this limit, stop the iterative process.
%
%OUTPUT
% xnew: the solution found by iterative method.
% k: actual number of iterations used to obtain the above solution.
% relError: array that contains the relative error found at each iteration
%
%example call
% A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
% b=[-4 11 -16 11 -4]; maxIter=200; errorLimit=0.0001; resLimit=0.00001;
%[x,k,relError]=nma_GaussSeidelIterativeSolver(A,...
%                            [1,1,1,1,1]',b',maxIter,errorLimit,resLimit)
%

%by Nasser Abbasi  3/26/07

% do some error checking on input....

if nargin ~=6
    error 'wrong number of arguments. 6 inputs are required';
end

TRUE=1; FALSE=0;

[res,msg]=nma_IterativeSolversIsValidInput(A,x,b,...
    maxIter,errorLimit,resLimit);
if ~res
    error(msg);
end

[nRow,nCol]=size(A);
xold=x(:);
b=b(:);
k=0;
relError=zeros(maxIter,1);

keepLookingForSolution=TRUE;

while keepLookingForSolution
    k=k+1;

    xnew=xold;
    for i=1:nRow
        xnew(i)=xnew(i)+ (b(i)-A(i,:)*xnew)/A(i,i);
    end

    currentError=norm(xnew-xold);
    relError(k)=currentError/norm(xnew);

    if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
        keepLookingForSolution=FALSE;
    else
        xold=xnew;
    end
end

end
```

4.9.3.5 nma_IterativeSolversIsValidInput.m

```
function [res,msg]=nma_IterativeSolversIsValidInput(A,x,b,maxIter,errorLimit,resLimit)
%function[res,msg]=nma_IterativeSolversIsValidInput(A,x,b,maxIter,errorLimit,resLimit)
%
%helper function. Called by iterative liner solvers to validate input
%
%Nasser Abbasi 03/26/07

res=0;
msg='';
if ~isnumeric(A)|~isnumeric(b)|~isnumeric(x)|~isnumeric(maxIter) ...
        |~isnumeric(errorLimit)|~isnumeric(resLimit)
    msg='non numeric input detected';
    return;
end

[nRow,nCol]=size(A);
if nRow~=nCol
    msg='square A matrix expected';
    return;
end

[m,n]=size(b);
if n>1
    msg='b must be a vector';
    return;
end

if m~=nRow
    msg='length of b does not match A matrix size';
    return;
end

[m,n]=size(x);
if n>1
    msg='x must be a vector';
    return;
end

if m~=nRow
    msg='length of x does not match A matrix size';
    return;
end

res=1;
return;

end
```

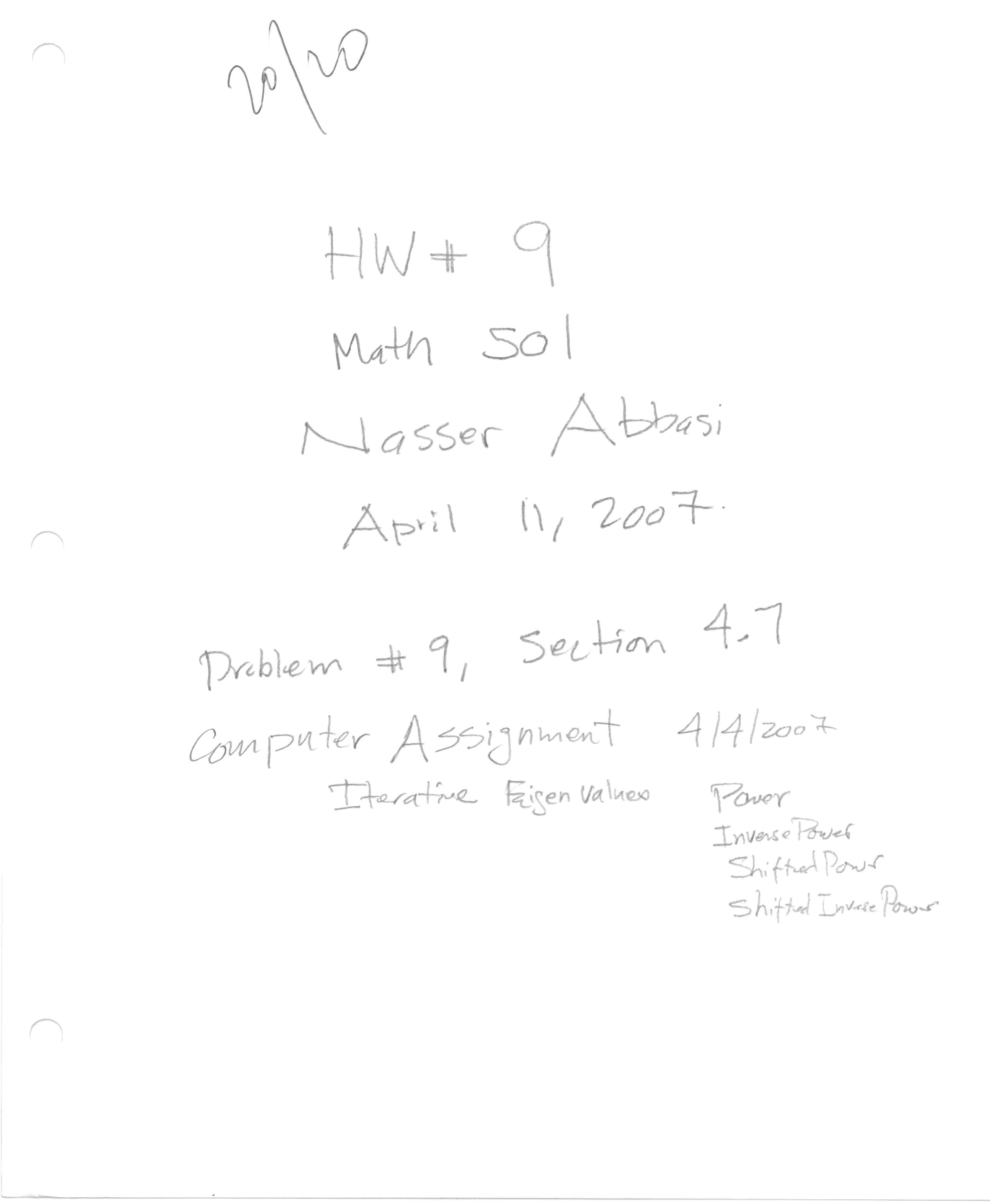
4.9.3.6 nma_SteepestIterativeSolver.m

```
function [xnew,k,relError]=nma_SteepestIterativeSolver(A,x,b,  ...
    maxIter,errorLimit,resLimit)
%function [xnew,k]=nma_SteepestIterativeSolver(A,x,b,...
%                                             maxIter,errorLimit,resLimit)
%
% Solve Ax=b using the Steepest descent Iterative method
%
%INPUT:
% A: the A matrix
% x: Initial guess for solution
% b: right hand side
% maxIter:  max number of iterations allowed
% errorLimit: error tolerance. difference between successive x iteration
%             values. if such a difference is less than this error, stop.
% resLimit: if |b-A*x| is less than this limit, stop the iterative process.
%
%OUTPUT
% xnew: the solution found by iterative method.
% k: actual number of iterations used to obtain the above solution.
% relError: array that contains the relative error found at each iteration
%
%example call
% A=[-4 2 1 0 0;1 -4 1 1 0;2 1 -4 1 2;0 1 1 -4 1;0 0 1 2 -4];
% b=[-4 11 -16 11 -4]; maxIter=200; errorLimit=0.0001; resLimit=0.00001;
%[x,k,relError]=nma_SteepestIterativeSolver(A,...
%                            [1,1,1,1,1]',b',maxIter,errorLimit,resLimit)
%

%by Nasser Abbasi  3/26/07

% do some error checking on input....

if nargin ~=6
    error 'wrong number of arguments. 6 inputs are required';
end

TRUE=1; FALSE=0;

[res,msg]=nma_IterativeSolversIsValidInput(A,x,b,...
    maxIter,errorLimit,resLimit);
if ~res
    error(msg);
end

[nRow,nCol]=size(A);
xold=x(:);
b=b(:);
k=0;
relError=zeros(maxIter,1);

keepLookingForSolution=TRUE;

while keepLookingForSolution
    k=k+1;

    v=b-A*xold;
    t=dot(v,v)/dot(v,A*v);
    xnew=xold+t*v;

    currentError=norm(xnew-xold);
    relError(k)=currentError/norm(xnew);

    if norm(b-A*xnew)<=resLimit || currentError<=errorLimit || k>maxIter
        keepLookingForSolution=FALSE;
    else
        xold=xnew;
    end
end

end
```

4.9.3.7 nma_driverTestSteepest.,

```
%
%This script is the driver to test steepest descent solver
%for computer assignment 3/19/07 for Math 501, CSUF
%
%Nasser Abbasi 033007
%
%file name: nma_driverTestSteepest.m

close all;
clear all;


maxIter=200;
errorLimit=0.0001;
resLimit=0.00001;

%%%% Now run another test. Use an SPD matrix, which is shown on
%page 245 of textbook (Numerical Analysis, Kincaid.Cheney)
fprintf('***** TEST steepest descent *********\n');
A=[10 1 2 3 4;1 9 -1 2 -3;2 -1 7 3 -5;3 2 3 12 -1;4 -3 -5 -1 15];
b=[12 -27 14 -17 12];
A
b
fprintf('======>Matlab linear solver solution, using A\\b  \n');
disp(A\b');

[x,k,relError]=nma_SteepestIterativeSolver(A,[1,1,1,1,1]',b', ...
    maxIter,errorLimit,resLimit);
fprintf('Steepest descent solution\n');
disp(x);
```

4.10 HW 9



20/20

HW # 9

Math 501

Nasser Abbasi

April 11, 2007.

Problem # 9, Section 4.7

Computer Assignment 4/4/2007

Iterative Eigen Values

Power

Inverse Power

Shifted Power

Shifted Inverse Power

problem # 9, section 4.7

let A be $n\times n$, and A-orthonormal system exist, show that A is SPD.

Solution outline express $\bar{x}$ in terms of basis $u^{(i)}$, start with definition of $\bar{x}^T A x$, show this must be >0 utilize property of A-orthonormal to cancel terms:

solution given A $n\times n$

given A-orthonormal system $[u^1|u^2 \cdots |u^n]$

i.e $\langle u^i, Au^j\rangle = \delta_{ij}$

i.e

$$(1)\quad \left\{ \begin{array}{lll} (u^1)^T A u^1 = 1 & (u^2)^T A u^1 = 0 & \cdots (u^n)^T A u^1 = 0 \\ (u^1)^T A u^2 = 0 & (u^2)^T A u^2 = 1 & (u^n)^T A u^2 = 0 \\ (u^1)^T A u^3 = 0, & \vdots, & \vdots \\ \vdots & & \\ (u^1)^T A u^n = 0 & (u^2)^T A u^n = 0 \cdots & (u^n)^T A u^n = 1 \end{array} \right.$$

now A is SPD if $\bar{x}^T A x > 0$ $(\bar{x} \neq 0)$

since u^i are Basis in $\mathbb{R}^n$, we can write vector $\bar{x}$ in terms of these basis. so

$$\bar{x} = \langle \bar{x}, \bar{u}^1\rangle \bar{u}^1 + \langle \bar{x}, \bar{u}^2\rangle \bar{u}^2 + \cdots + \langle \bar{x}, \bar{u}^n\rangle \bar{u}^n$$

hence using ~~this~~, we now write $\bar{x}^T A \bar{x}$ and compare to (1)

$$\bar{x}^T A \bar{x} = \left(\langle \bar{x}, \bar{u}^1\rangle \bar{u}^1 + \langle \bar{x}, \bar{u}^2\rangle \bar{u}^2 + \cdots + \langle \bar{x}, \bar{u}^n\rangle \bar{u}^n\right)^T A \left(\langle \bar{x}, \bar{u}^1\rangle \bar{u}^1 + + \langle \bar{x}, \bar{u}^n\rangle u^n\right)$$

let $\langle \bar{x}, \bar{u}^1\rangle = a_1$, $\langle \bar{x}, \bar{u}^2\rangle = a_2$ etc... these are the coordinates of $\bar{x}$ in this subspace. so we write

$$\bar{x}^T A \bar{x} = \left(a_1 \bar{u}^1 + a_2 \bar{u}^2 + \cdots + a_n \bar{u}^n\right)^T A \left(a_1 \bar{u}^1 + a_2 \bar{u}^2 + \cdots + a_n \bar{u}^n\right)$$

$$= \left(a_1 \bar{u}^1 + a_2 \bar{u}^2 + \cdots + a_n \bar{u}^n\right)^T \left(A a_1 \bar{u}^1 + A a_2 \bar{u}^2 + \cdots + A a_n \bar{u}^n\right)$$

$$= a_1^2 (\bar{u}^1)^T A u_1^1 + a_1 a_2 \bar{u}^{1T} A \bar{u}^2 + a_1 a_3 \bar{u}^{1T} A \bar{u}^3 + \cdots + a_1 a_n (\bar{u}^1)^T A \bar{u}^n$$
$$+ a_2 a_1 (\bar{u}^2)^T A u^1 + a_2^2 (\bar{u}^2)^T A \bar{u}^2 + a_2 a_3 (u^2)^T A \bar{u}^3 + \cdots + a_2 a_n (\bar{u}^2)^T A \bar{u}^n$$
$$\cdots + \cdots + a_n a_1 (\bar{u}^n)^T A u^1 + a_n a_2 (\bar{u}^n)^T A \bar{u}^2 + \cdots + \quad + a_n^2 (\bar{u}^n)^T A \bar{u}^T \longrightarrow$$

So we see the pattern for $\bar{x}^T A \bar{x}$ as

$$= a_1^2 (\bar{u}^1)^T A \bar{u}^1 + \cdots + \cdots \qquad (2)$$

$$+ a_2^2 (\bar{u}^2)^T A \bar{u}^2 + \cdots$$

$$\cdots \qquad + a_3^2 (\bar{u}^3)^T A \bar{u}^3 + \cdots$$

But since from (1) we see that $(u^1)^T A \bar{u}^1 = 1$, $(u^2)^T A \bar{u}^2 = 1, \ldots, (u^n)^T A u^n = 1$ and everything else is <u>zero</u>, then (2) can be written as

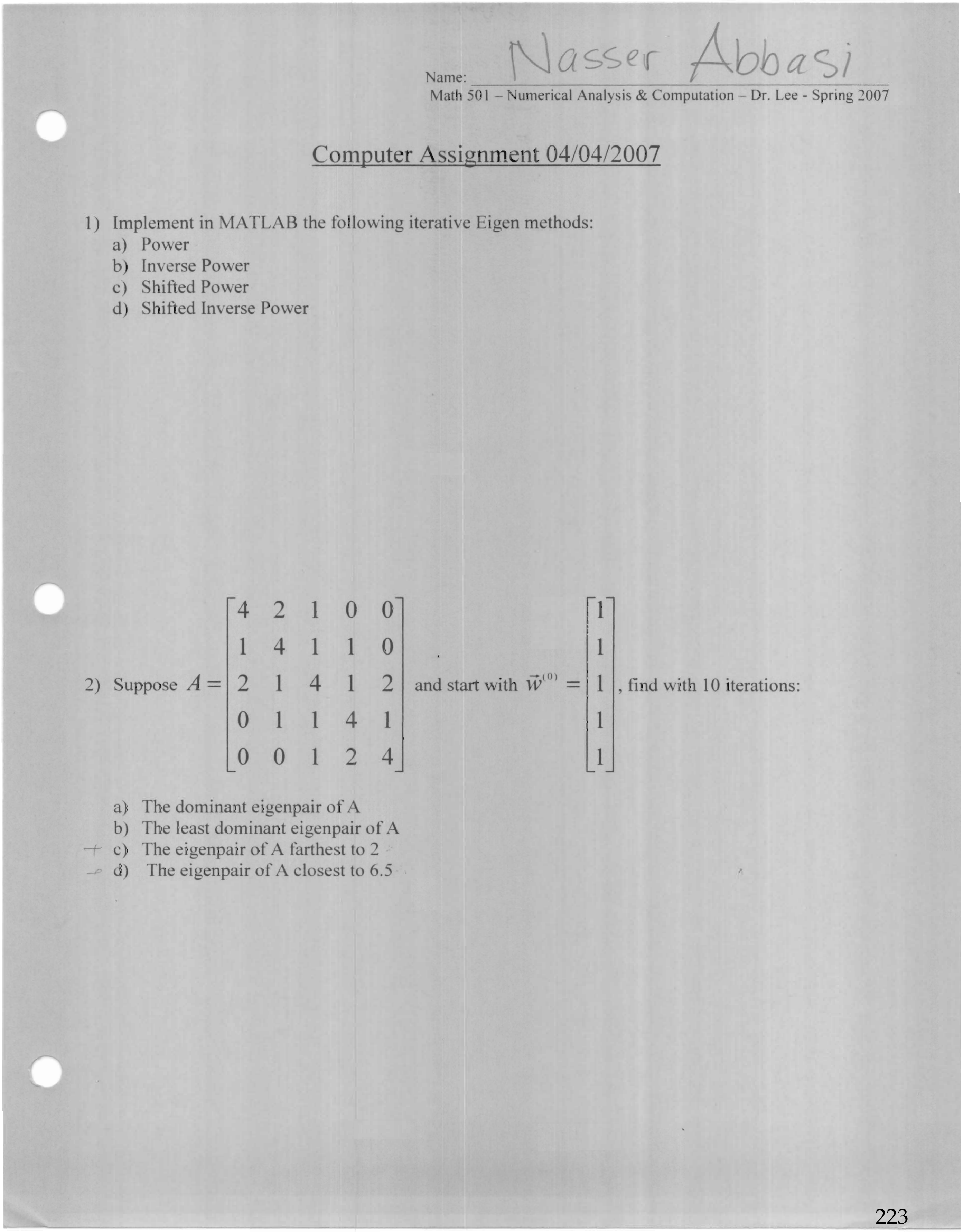
$$\bar{x}^T A x = a_1^2 + a_2^2 + a_3^2 + \cdots + a_n^2$$

but the a_i are the coordinates of $\bar{x}$ in R^n. and for $\bar{x} \neq \bar{0}$, then at least <u>one</u> coordinate must not vanish. hence $\bar{x}^T A x \neq 0$. in addition, since the coordinates are all <u>squared</u> hence it is positive sum.

10/10

$$\Longrightarrow \boxed{\bar{x}^T A x > 0}$$ ie A is SPD

QED



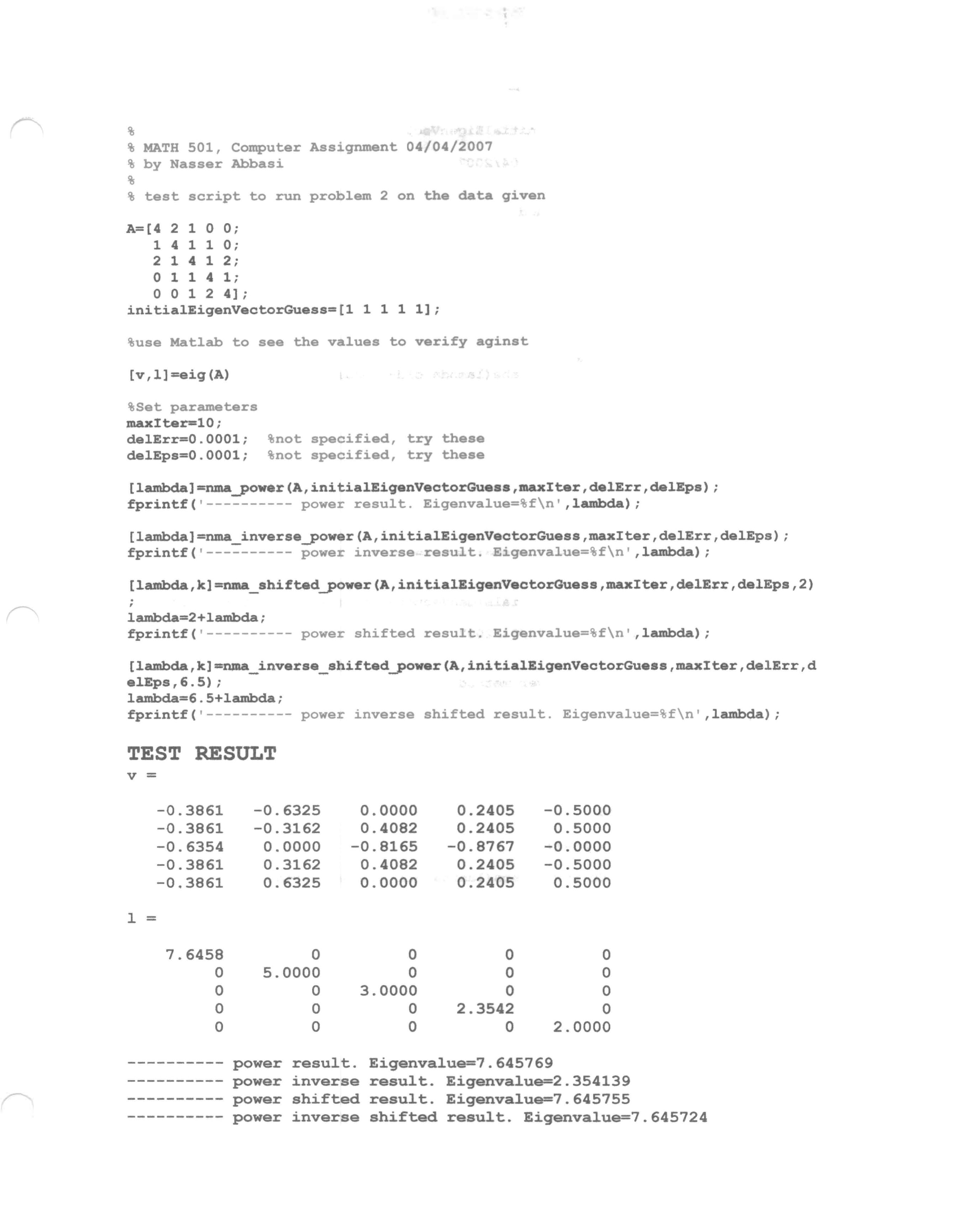
Name: Nasser Abbasi

Math 501 – Numerical Analysis & Computation – Dr. Lee - Spring 2007

Computer Assignment 04/04/2007

1) Implement in MATLAB the following iterative Eigen methods:
   a) Power
   b) Inverse Power
   c) Shifted Power
   d) Shifted Inverse Power

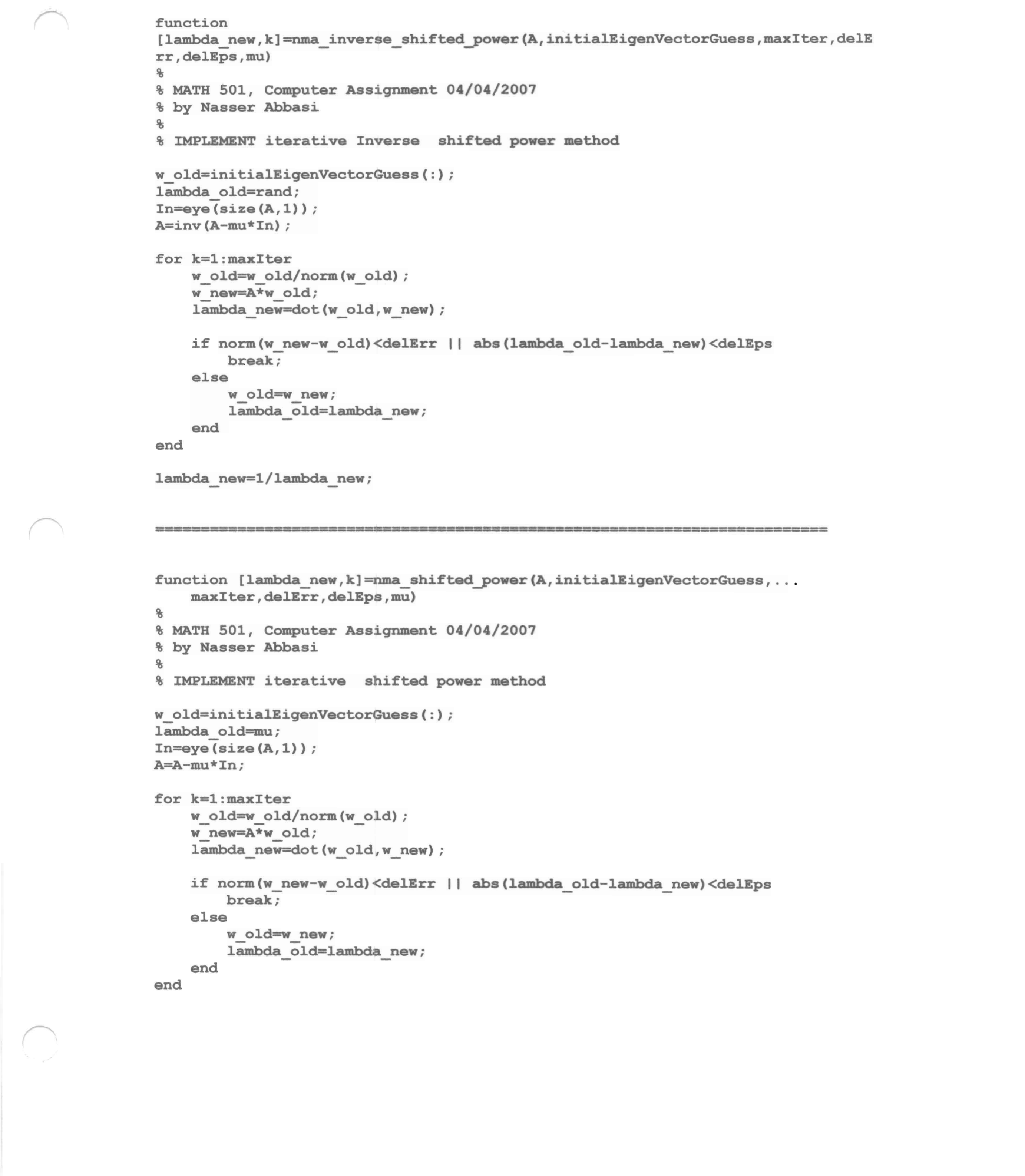
2) Suppose $A = \begin{bmatrix} 4 & 2 & 1 & 0 & 0 \\ 1 & 4 & 1 & 1 & 0 \\ 2 & 1 & 4 & 1 & 2 \\ 0 & 1 & 1 & 4 & 1 \\ 0 & 0 & 1 & 2 & 4 \end{bmatrix}$ and start with $\vec{w}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$, find with 10 iterations:
   a) The dominant eigenpair of A
   b) The least dominant eigenpair of A
   c) The eigenpair of A farthest to 2
   d) The eigenpair of A closest to 6.5



```
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% test script to run problem 2 on the data given

A=[4 2 1 0 0;
   1 4 1 1 0;
   2 1 4 1 2;
   0 1 1 4 1;
   0 0 1 2 4];
initialEigenVectorGuess=[1 1 1 1 1];

%use Matlab to see the values to verify aginst

[v,l]=eig(A)

%Set parameters
maxIter=10;
delErr=0.0001;  %not specified, try these
delEps=0.0001;  %not specified, try these

[lambda]=nma_power(A,initialEigenVectorGuess,maxIter,delErr,delEps);
fprintf('---------- power result. Eigenvalue=%f\n',lambda);

[lambda]=nma_inverse_power(A,initialEigenVectorGuess,maxIter,delErr,delEps);
fprintf('---------- power inverse result. Eigenvalue=%f\n',lambda);

[lambda,k]=nma_shifted_power(A,initialEigenVectorGuess,maxIter,delErr,delEps,2)
;
lambda=2+lambda;
fprintf('---------- power shifted result. Eigenvalue=%f\n',lambda);

[lambda,k]=nma_inverse_shifted_power(A,initialEigenVectorGuess,maxIter,delErr,d
elEps,6.5);
lambda=6.5+lambda;
fprintf('---------- power inverse shifted result. Eigenvalue=%f\n',lambda);
```

TEST RESULT

```
v =

   -0.3861   -0.6325    0.0000    0.2405   -0.5000
   -0.3861   -0.3162    0.4082    0.2405    0.5000
   -0.6354    0.0000   -0.8165   -0.8767   -0.0000
   -0.3861    0.3162    0.4082    0.2405   -0.5000
   -0.3861    0.6325    0.0000    0.2405    0.5000


l =

    7.6458         0         0         0         0
         0    5.0000         0         0         0
         0         0    3.0000         0         0
         0         0         0    2.3542         0
         0         0         0         0    2.0000

---------- power result. Eigenvalue=7.645769
---------- power inverse result. Eigenvalue=2.354139
---------- power shifted result. Eigenvalue=7.645755
---------- power inverse shifted result. Eigenvalue=7.645724
```



```
function
[lambda_new,k]=nma_inverse_shifted_power(A,initialEigenVectorGuess,maxIter,delE
rr,delEps,mu)
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative Inverse  shifted power method

w_old=initialEigenVectorGuess(:);
lambda_old=rand;
In=eye(size(A,1));
A=inv(A-mu*In);

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);

    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end

lambda_new=1/lambda_new;


===========================================================================


function [lambda_new,k]=nma_shifted_power(A,initialEigenVectorGuess,...
    maxIter,delErr,delEps,mu)
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative  shifted power method

w_old=initialEigenVectorGuess(:);
lambda_old=mu;
In=eye(size(A,1));
A=A-mu*In;

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);

    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end
```

4.10.1 Source code

4.10.1.1 nma_inverse_power.m

```
nction lambda_new=nma_inverse_power(A,initialEigenVectorGuess,maxIter,delErr,delEps)
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative Inverse power method

A=inv(A);

w_old=initialEigenVectorGuess(:);
lambda_old=rand;

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);

    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end

lambda_new=1/lambda_new;
```

4.10.1.2 nma_inverse_shifted_power.m

```
function [lambda_new,k]=nma_inverse_shifted_power(A,initialEigenVectorGuess,maxIter,delErr,del
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative Inverse  shifted power method

w_old=initialEigenVectorGuess(:);
lambda_old=rand;
In=eye(size(A,1));
A=inv(A-mu*In);

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);

    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end

lambda_new=1/lambda_new;
```

4.10.1.3 nma_power.m

```
function lambda_new=nma_power(A,initialEigenVectorGuess,maxIter,delErr,delEps)
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative  power method

w_old=initialEigenVectorGuess(:);

lambda_old=rand;

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);
    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end
```

4.10.1.4 nma_shifted_power.m

```
nction [lambda_new,k]=nma_shifted_power(A,initialEigenVectorGuess, ...
    maxIter,delErr,delEps,mu)
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% IMPLEMENT iterative  shifted power method
```

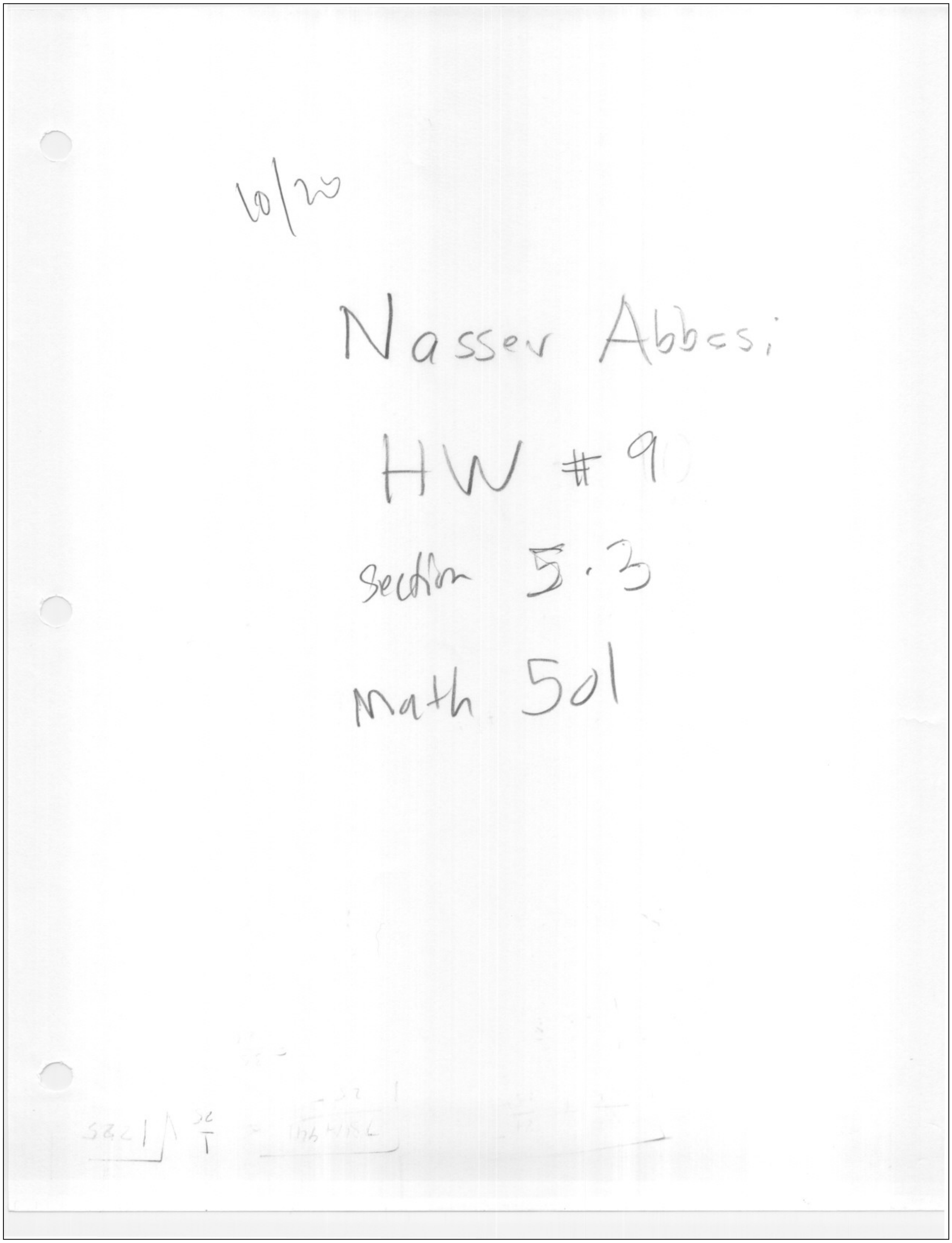
```
w_old=initialEigenVectorGuess(:);
lambda_old=mu;
In=eye(size(A,1));
A=A-mu*In;

for k=1:maxIter
    w_old=w_old/norm(w_old);
    w_new=A*w_old;
    lambda_new=dot(w_old,w_new);

    if norm(w_new-w_old)<delErr || abs(lambda_old-lambda_new)<delEps
        break;
    else
        w_old=w_new;
        lambda_old=lambda_new;
    end
end
```

4.10.1.5 nma_test.m

```
%
% MATH 501, Computer Assignment 04/04/2007
% by Nasser Abbasi
%
% test script to run problem 2 on the data given

A=[4 2 1 0 0;
   1 4 1 1 0;
   2 1 4 1 2;
   0 1 1 4 1;
   0 0 1 2 4];
initialEigenVectorGuess=[1 1 1 1 1];

%use Matlab to see the values to verify aginst

[v,l]=eig(A)

%Set parameters
maxIter=10;
delErr=0.0001;  %not specified, try these
delEps=0.0001;  %not specified, try these

[lambda]=nma_power(A,initialEigenVectorGuess,maxIter,delErr,delEps);
fprintf('---------- power result. Eigenvalue=%f\n',lambda);

[lambda]=nma_inverse_power(A,initialEigenVectorGuess,maxIter,delErr,delEps);
fprintf('---------- power inverse result. Eigenvalue=%f\n',lambda);
```

```
[lambda,k]=nma_shifted_power(A,initialEigenVectorGuess,maxIter,delErr,delEps,2);
lambda=2+lambda;
fprintf('---------- power shifted result. Eigenvalue=%f\n',lambda);

[lambda,k]=nma_inverse_shifted_power(A,initialEigenVectorGuess,maxIter,delErr,delEps,6.5);
lambda=6.5+lambda;
fprintf('---------- power inverse shifted result. Eigenvalue=%f\n',lambda);
```

4.11 HW 10



10/20

Nasser Abbasi

HW # 9

section 5.3

math 501

5.3 # 2

Prove that if A is $m \times n$ matrix of rank n, then $A^{\#}A$ is nonsingular

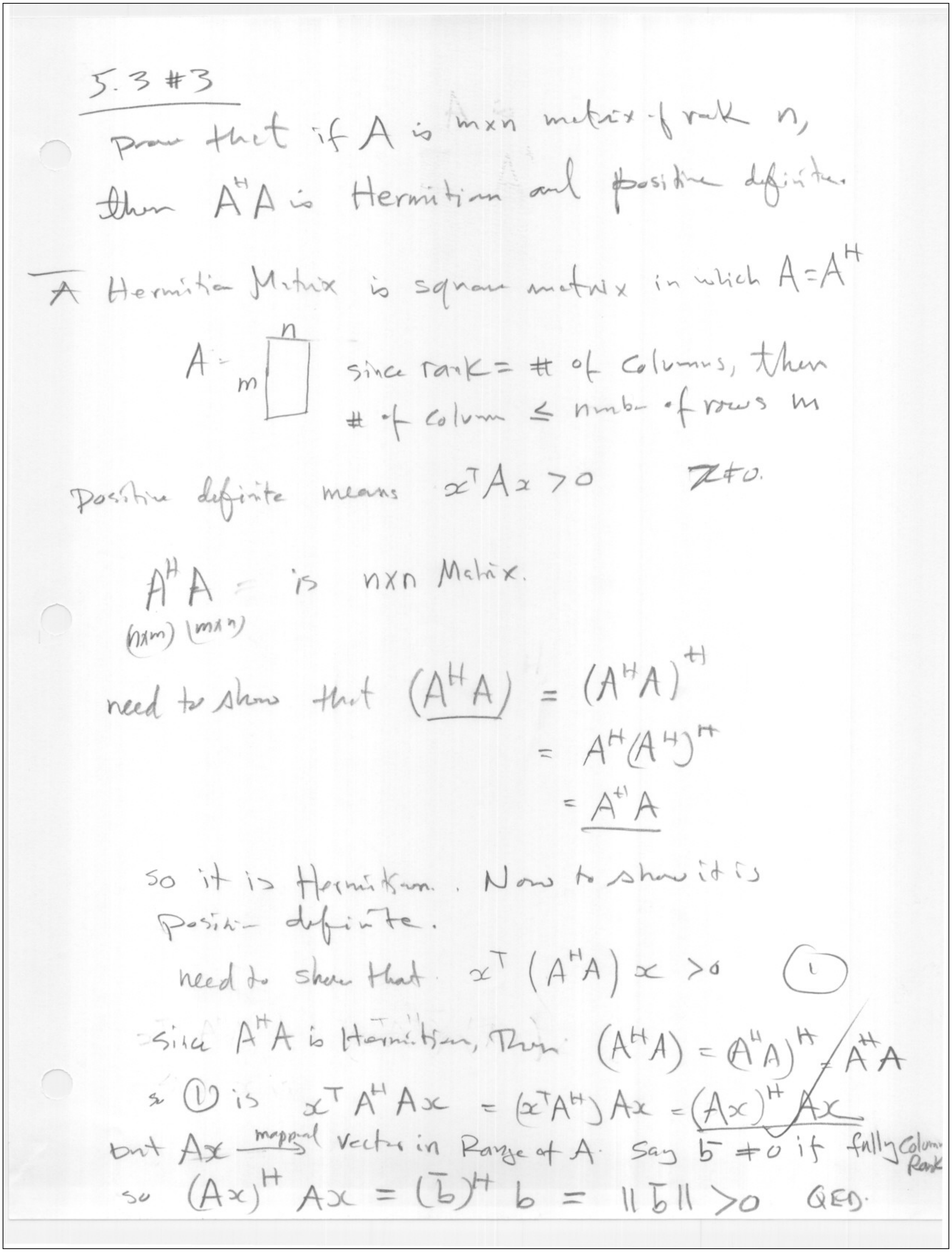
, $\underset{(n \times m)}{A^{H}}\underset{(m \times n)}{A} =$ is matrix that square of order $n \times n$.

need to show $A^{\#}A$ Rows

but since $\underline{\text{Rank} = n}$, then

We have n L.I. columns, and n L.I rows. So Matrix is full column Rank and full Row Rank.

$\Longrightarrow$ Non Singular.



5.3 #3

prove that if A is m×n matrix of rank n, then $A^H A$ is Hermitian and positive definite.

A Hermitian Matrix is square matrix in which $A=A^H$

$A = $ (n, m) since rank = # of columns, then # of column $\leq$ number of rows m

Positive definite means $x^T A x > 0$ $x \neq 0$.

$\underset{(n\times m)}{A^H}\ \underset{(m\times n)}{A}$ is n×n Matrix.

need to show that

$$\underline{(A^H A)} = (A^H A)^H = A^H (A^H)^H = \underline{A^H A}$$

so it is Hermitian. Now to show it is positive definite.

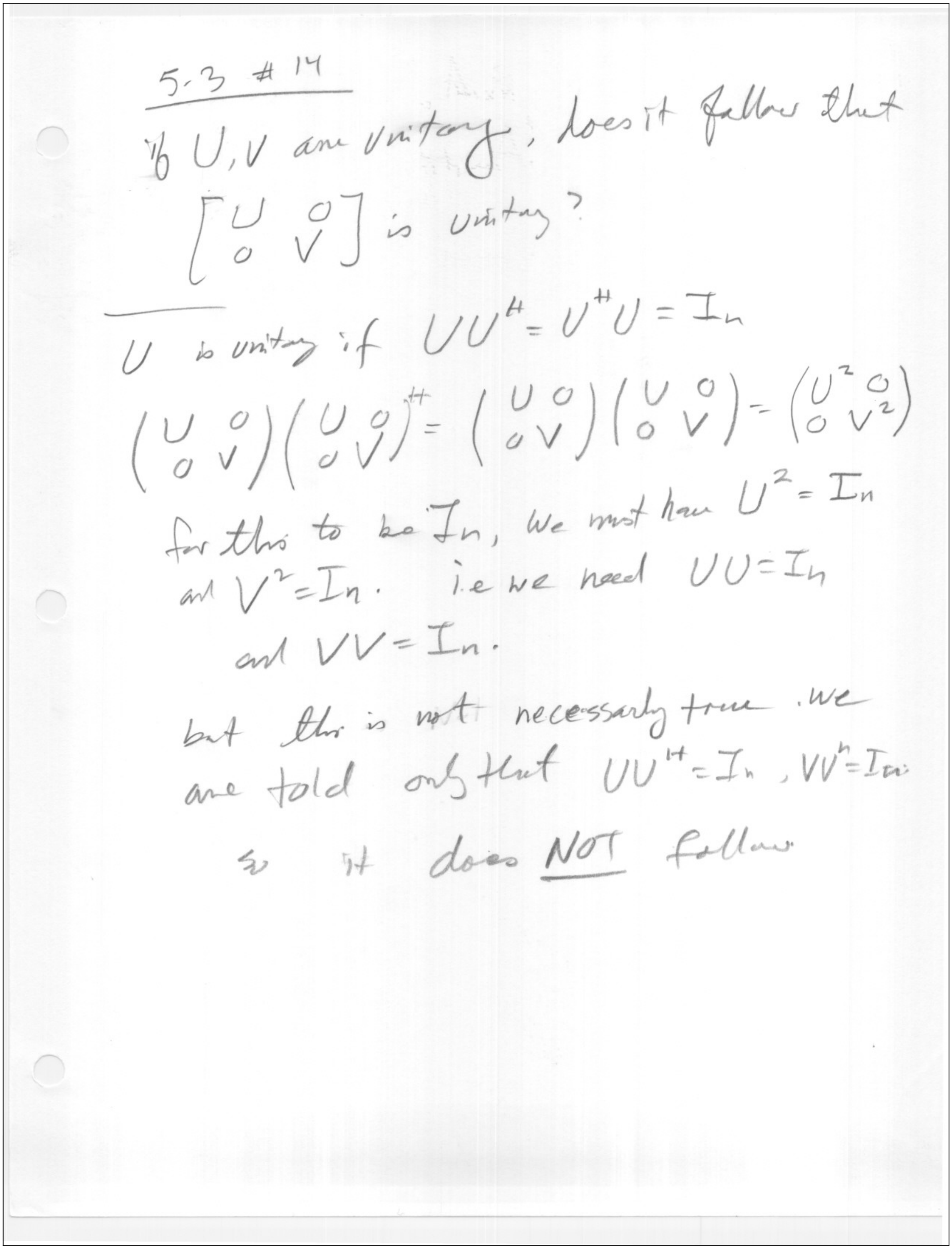
need to show that $x^T (A^H A) x > 0$ (1)

since $A^H A$ is Hermitian, Then $(A^H A) = (A^H A)^H = A^H A$

so (1) is $x^T A^H A x = (x^T A^H) A x = \underline{(Ax)^H Ax}$

but $Ax \xrightarrow{\text{mapped}}$ vector in Range of A. Say $\bar{b} \neq 0$ if fully Column Rank

so $(Ax)^H Ax = (\bar{b})^H b = \|\bar{b}\| > 0$ QED.



5.3 # 14

If U, V are unitary, does it follow that $\begin{bmatrix} U & 0 \\ 0 & V \end{bmatrix}$ is unitary?

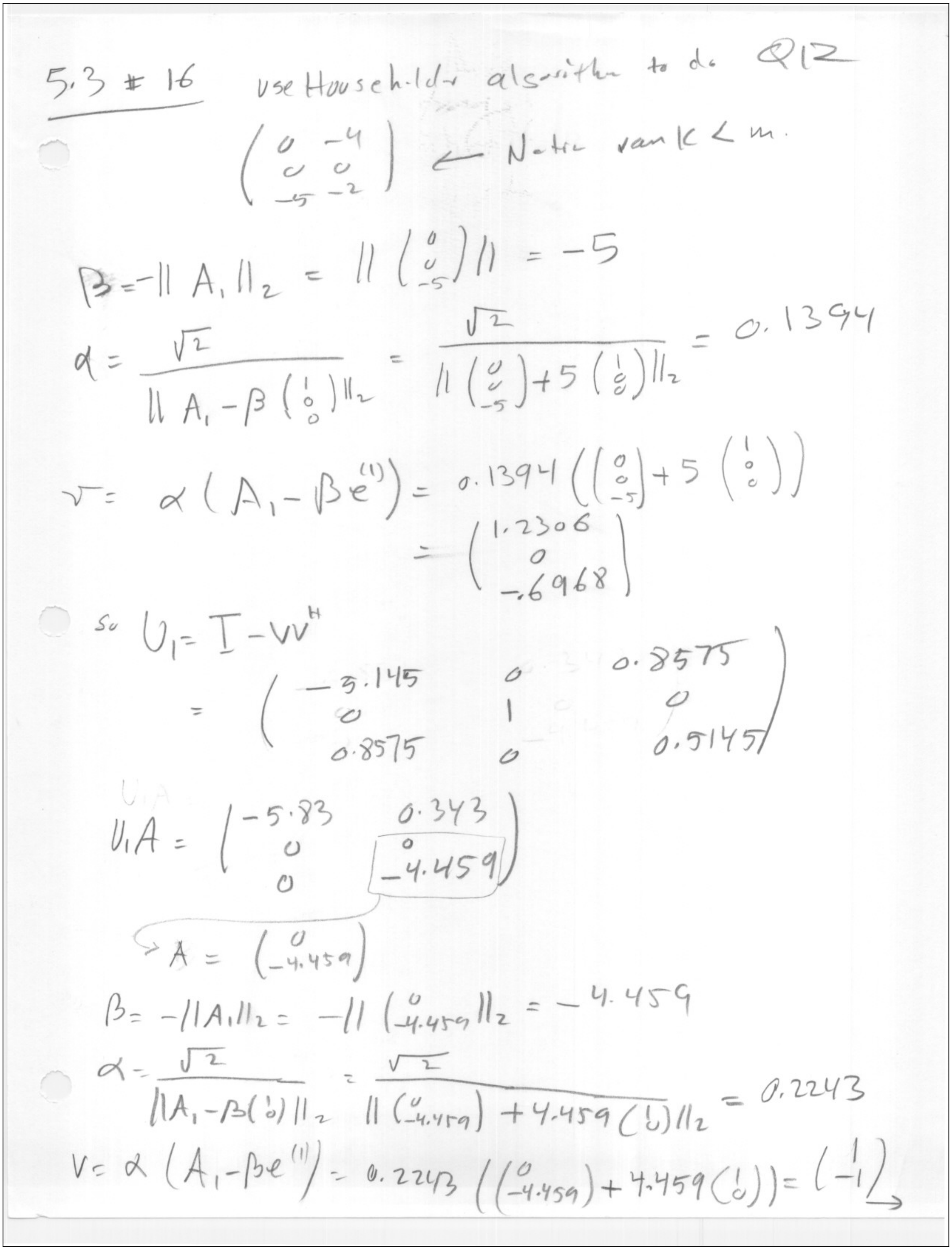
U is unitary if $UU^H = U^H U = I_n$

$$\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}^H = \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} \begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix} = \begin{pmatrix} U^2 & 0 \\ 0 & V^2 \end{pmatrix}$$

for this to be I_n, we must have $U^2 = I_n$ and $V^2 = I_n$. i.e we need $UU = I_n$ and $VV = I_n$.

but this is not necessarly true. we are told only that $UU^H = I_n$, $VV^h = I_n$.

so it does NOT follow



5.3 # 16 use Householder algorithm to do QR

$$\begin{pmatrix}0 & -4\\ 0 & 0\\ -5 & -2\end{pmatrix} \leftarrow \text{Notice rank} < m.$$

$$\beta = -\|A_1\|_2 = \left\|\begin{pmatrix}0\\0\\-5\end{pmatrix}\right\| = -5$$

$$\alpha = \frac{\sqrt{2}}{\|A_1 - \beta\begin{pmatrix}1\\0\\0\end{pmatrix}\|_2} = \frac{\sqrt{2}}{\|\begin{pmatrix}0\\0\\-5\end{pmatrix} + 5\begin{pmatrix}1\\0\\0\end{pmatrix}\|_2} = 0.1394$$

$$v = \alpha\left(A_1 - \beta e^{(1)}\right) = 0.1394\left(\begin{pmatrix}0\\0\\-5\end{pmatrix} + 5\begin{pmatrix}1\\0\\0\end{pmatrix}\right)$$
$$= \begin{pmatrix}1.2306\\0\\-.6968\end{pmatrix}$$

so $U_1 = I - vv^H$
$$= \begin{pmatrix}-5.145 & 0 & 0.8575\\ 0 & 1 & 0\\ 0.8575 & 0 & 0.5145\end{pmatrix}$$

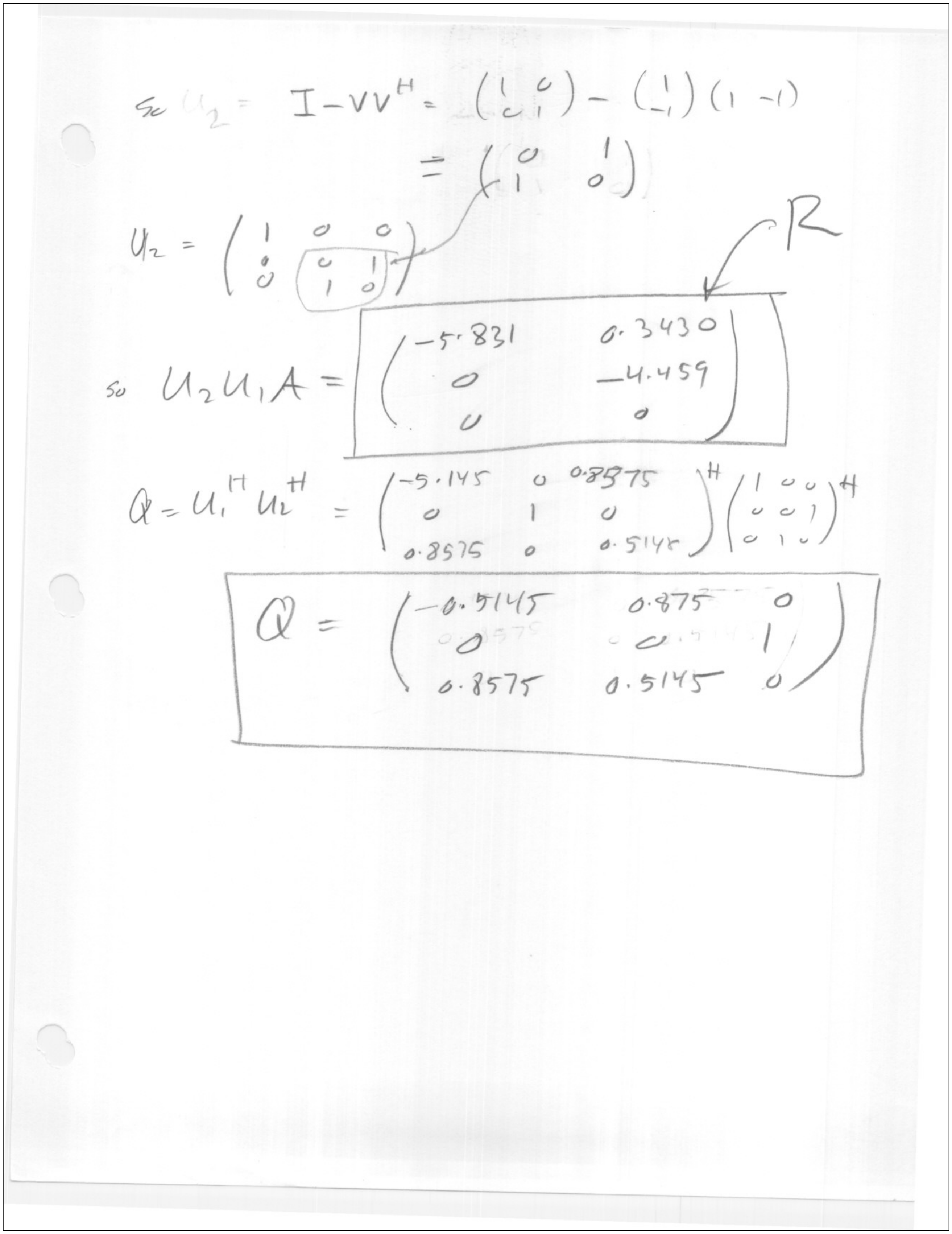
$$U_1A = \begin{pmatrix}-5.83 & 0.343\\ 0 & \boxed{\begin{matrix}0\\-4.459\end{matrix}}\\ 0 & \end{pmatrix}$$

$$\rightarrow A = \begin{pmatrix}0\\-4.459\end{pmatrix}$$

$$\beta = -\|A_1\|_2 = -\left\|\begin{pmatrix}0\\-4.459\end{pmatrix}\right\|_2 = -4.459$$

$$\alpha = \frac{\sqrt{2}}{\|A_1 - \beta\begin{pmatrix}1\\0\end{pmatrix}\|_2} = \frac{\sqrt{2}}{\|\begin{pmatrix}0\\-4.459\end{pmatrix} + 4.459\begin{pmatrix}1\\0\end{pmatrix}\|_2} = 0.2243$$

$$v = \alpha\left(A_1 - \beta e^{(1)}\right) = 0.2243\left(\begin{pmatrix}0\\-4.459\end{pmatrix} + 4.459\begin{pmatrix}1\\0\end{pmatrix}\right) = \begin{pmatrix}1\\-1\end{pmatrix} \rightarrow$$



$$\text{So } U_2 = I - vv^H = \begin{pmatrix}1&0\\0&1\end{pmatrix} - \begin{pmatrix}1\\-1\end{pmatrix}\begin{pmatrix}1 & -1\end{pmatrix}$$

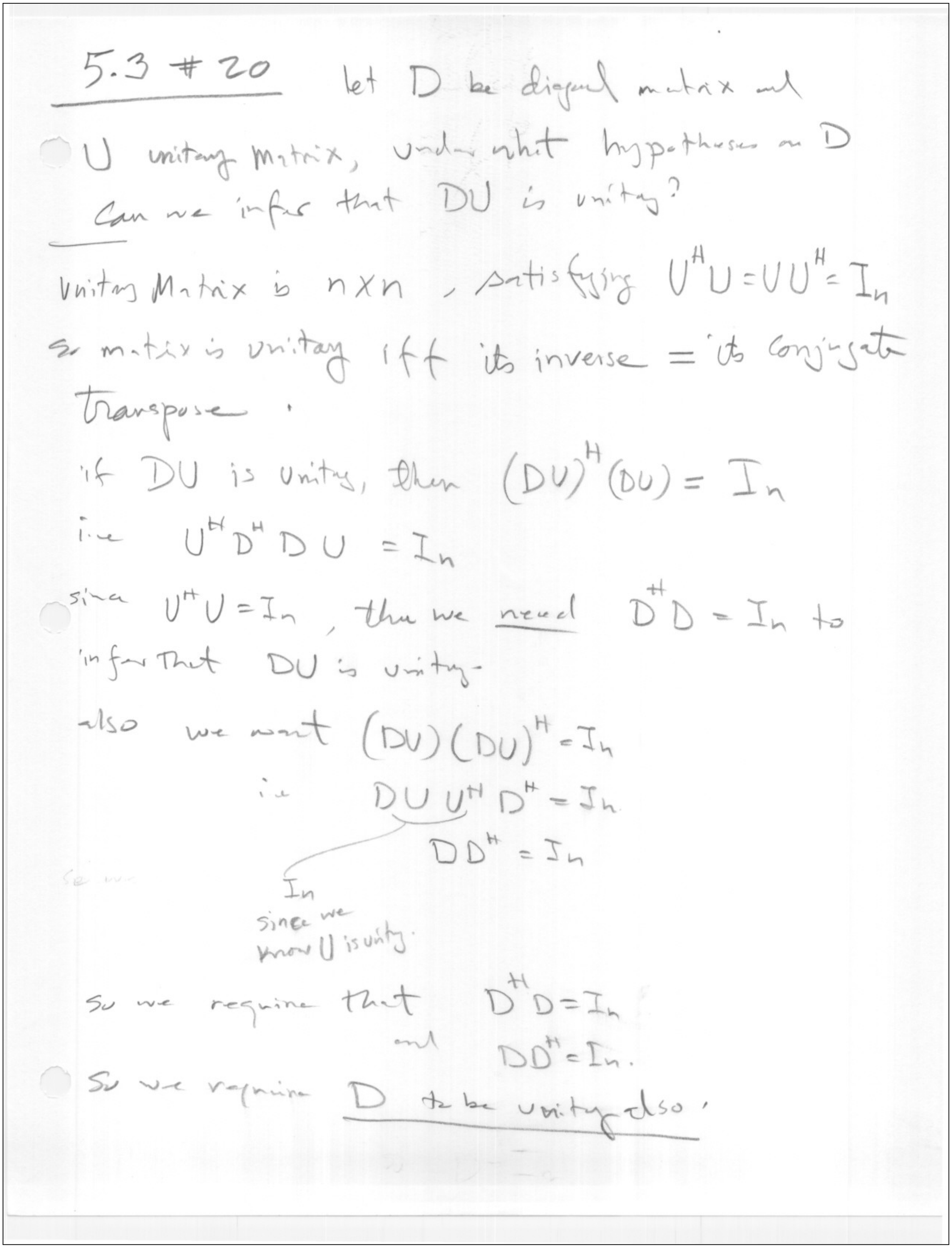
$$= \begin{pmatrix}0&1\\1&0\end{pmatrix}$$

$$U_2 = \begin{pmatrix}1 & \begin{matrix}0&0\end{matrix}\\ \begin{matrix}0\\0\end{matrix} & \boxed{\begin{matrix}0&1\\1&0\end{matrix}}\end{pmatrix}$$

$$\text{so } U_2 U_1 A = \overset{R}{\boxed{\begin{pmatrix}-5.831 & 0.3430\\ 0 & -4.459\\ 0 & 0\end{pmatrix}}}$$

$$Q = U_1^H U_2^H = \begin{pmatrix}-5.145 & 0 & 0.8575\\ 0 & 1 & 0\\ 0.8575 & 0 & 0.5145\end{pmatrix}^H \begin{pmatrix}1&0&0\\0&0&1\\0&1&0\end{pmatrix}^H$$

$$\boxed{Q = \begin{pmatrix}-0.5145 & 0.875 & 0\\ 0 & 0 & 1\\ 0.8575 & 0.5145 & 0\end{pmatrix}}$$



5.3 #20 let D be diagonal matrix and U unitary matrix, under what hypotheses on D can we infer that DU is unitary?

Unitary Matrix is $n \times n$, satisfying $U^H U = U U^H = I_n$

so matrix is unitary iff its inverse = its conjugate transpose.

if DU is unitary, then $(DU)^H (DU) = I_n$

i.e $U^H D^H D U = I_n$

since $U^H U = I_n$, then we need $D^H D = I_n$ to infer that DU is unitary.

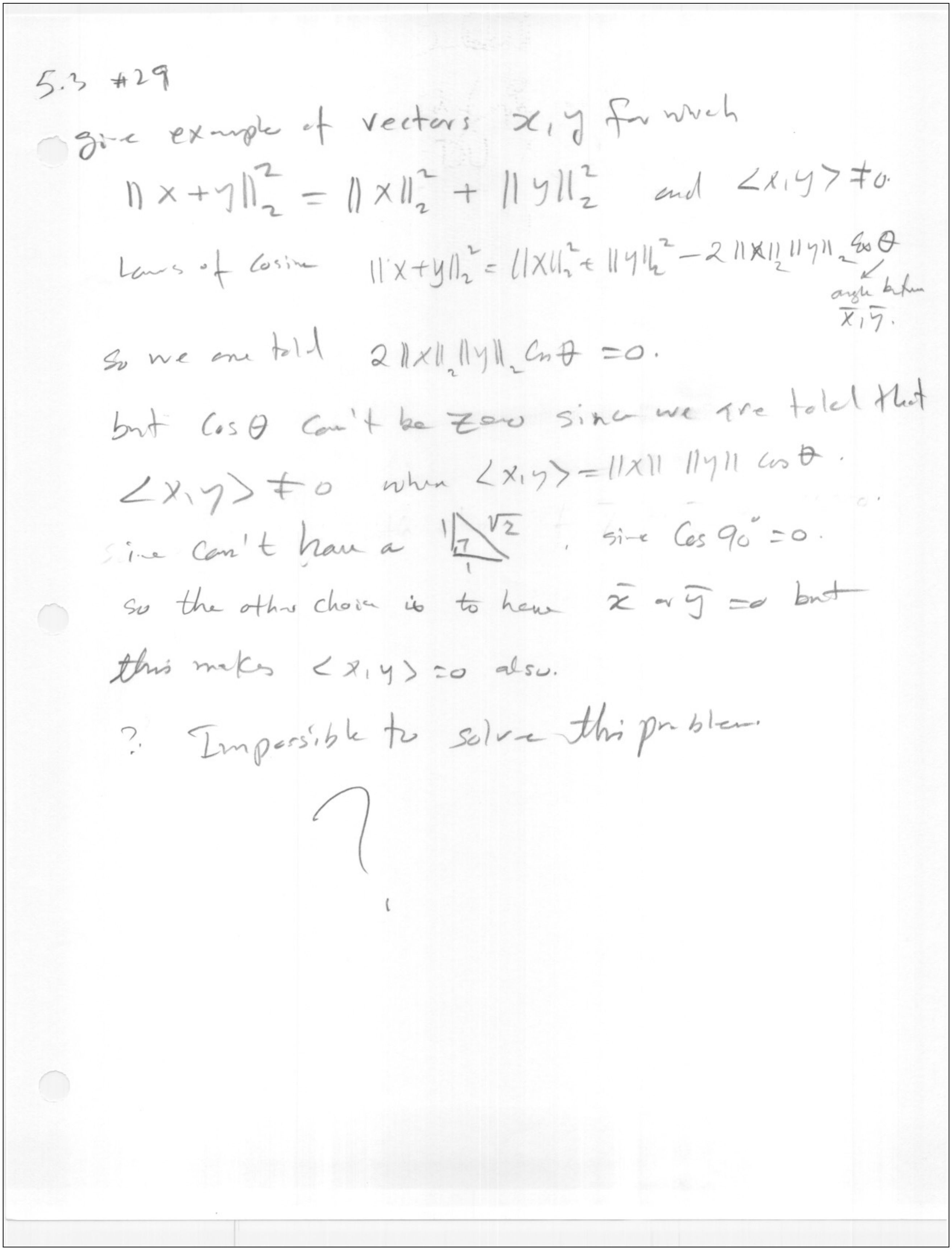
also we want $(DU)(DU)^H = I_n$

i.e $D \underbrace{U U^H}_{\substack{I_n \\ \text{since we} \\ \text{know } U \text{ is unitary.}}} D^H = I_n$

$DD^H = I_n$

So we require that $D^H D = I_n$ and $DD^H = I_n$.

So we require D to be unitary also.



5.3 #29

give example of vectors x, y for which

$$\|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 \quad \text{and} \quad \langle x,y\rangle \neq 0.$$

Laws of Cosine $\|x+y\|_2^2 = \|x\|_2^2 + \|y\|_2^2 - 2\|x\|_2\|y\|_2 \cos\theta$

angle between $\bar{x}, \bar{y}$.

So we are told $2\|x\|_2\|y\|_2 \cos\theta = 0$.

but $\cos\theta$ can't be zero since we are told that

$\langle x,y\rangle \neq 0$ when $\langle x,y\rangle = \|x\| \, \|y\| \cos\theta$.

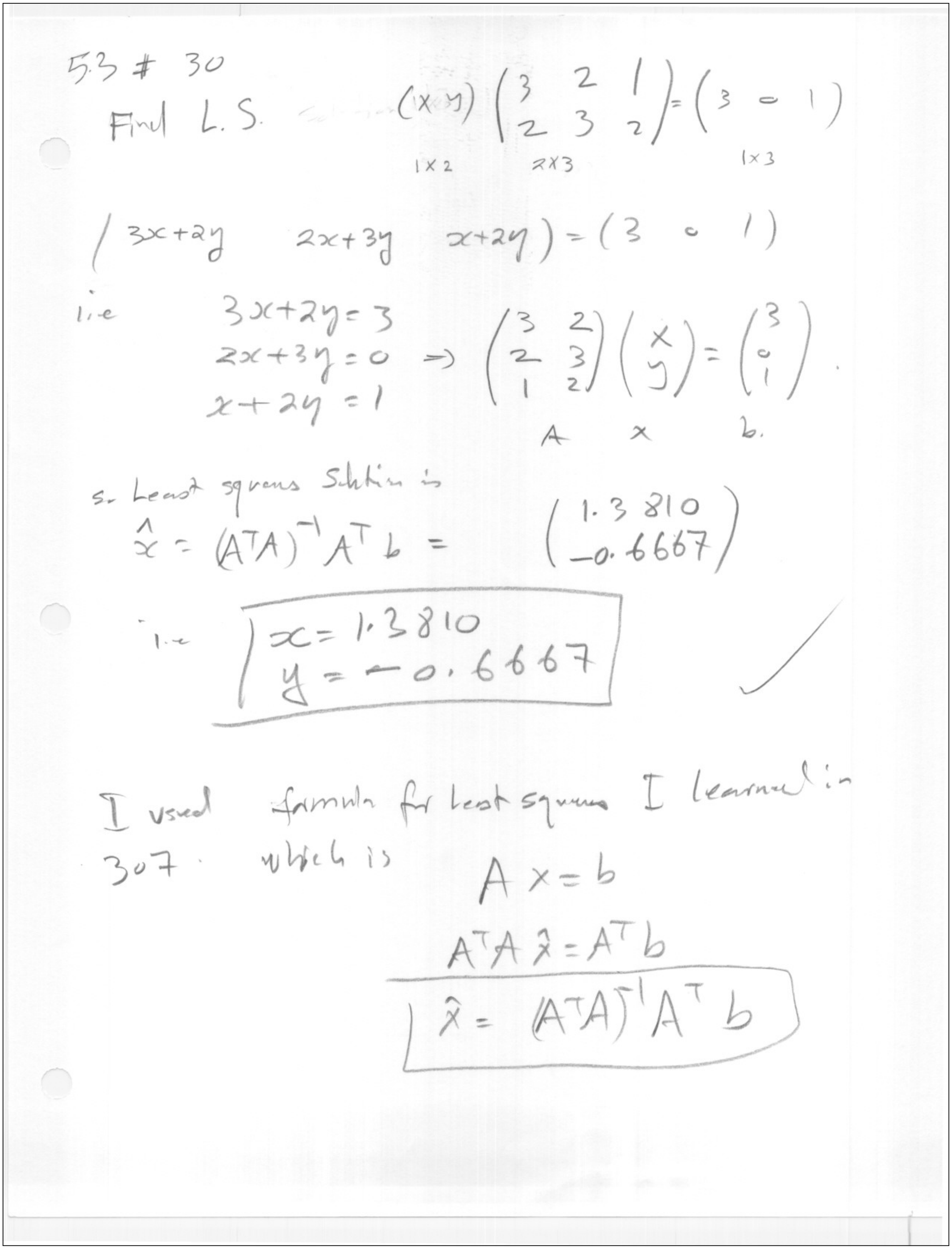
Since can't have a 1 $\sqrt{2}$ 1, since $\cos 90^\circ = 0$.

so the other choice is to have $\bar{x}$ or $\bar{y} = 0$ but

this makes $\langle x,y\rangle = 0$ also.

$\therefore$ Impossible to solve this problem.

?



5.3 # 30

Find L.S. $(x\ y)\begin{pmatrix} 3 & 2 & 1 \\ 2 & 3 & 2 \end{pmatrix} = (3\ 0\ 1)$

1×2 $\quad 2\times 3 \quad$ 1×3

$(3x+2y \quad 2x+3y \quad x+2y) = (3 \quad 0 \quad 1)$

i.e $\begin{aligned} 3x+2y&=3 \\ 2x+3y&=0 \\ x+2y&=1 \end{aligned} \Rightarrow \underbrace{\begin{pmatrix} 3 & 2 \\ 2 & 3 \\ 1 & 2 \end{pmatrix}}_{A} \underbrace{\begin{pmatrix} x \\ y \end{pmatrix}}_{x} = \underbrace{\begin{pmatrix} 3 \\ 0 \\ 1 \end{pmatrix}}_{b.}$.

so Least squares Solution is

$\hat{x} = (A^TA)^{-1}A^T b = \begin{pmatrix} 1.3810 \\ -0.6667 \end{pmatrix}$

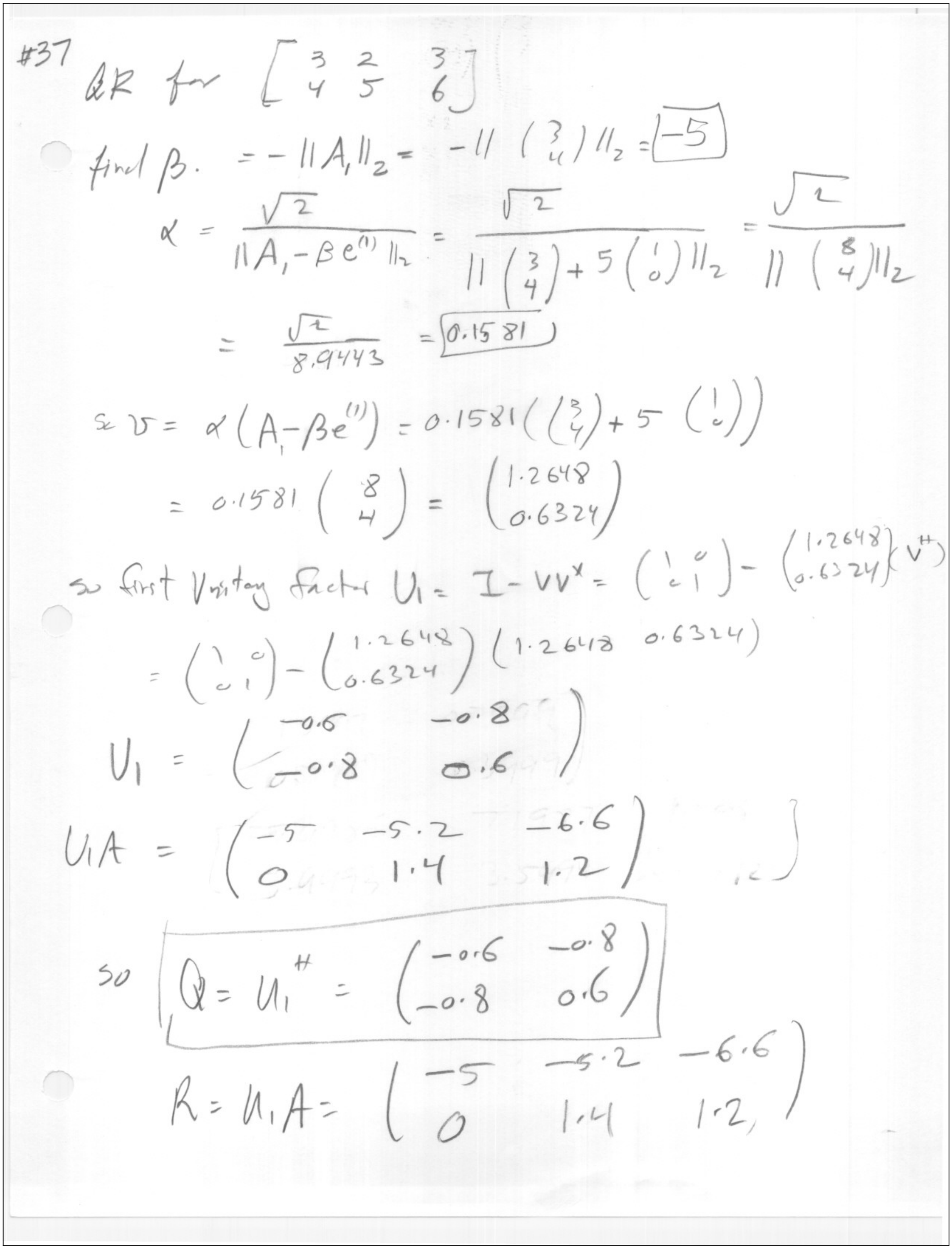
i.e $\boxed{\begin{aligned} x &= 1.3810 \\ y &= -0.6667 \end{aligned}}$ ✓

I used formula for least squares I learned in 307. which is

$$A x = b$$

$$A^TA\hat{x} = A^Tb$$

$$\boxed{\hat{x} = (A^TA)^{-1}A^T b}$$



#37 QR for $\begin{bmatrix}3&2&3\\4&5&6\end{bmatrix}$

find β. $= -\|A_1\|_2 = -\|\begin{pmatrix}3\\4\end{pmatrix}\|_2 = \boxed{-5}$

$$\alpha = \frac{\sqrt{2}}{\|A_1-\beta e^{(1)}\|_2} = \frac{\sqrt{2}}{\|\begin{pmatrix}3\\4\end{pmatrix}+5\begin{pmatrix}1\\0\end{pmatrix}\|_2} = \frac{\sqrt{2}}{\|\begin{pmatrix}8\\4\end{pmatrix}\|_2}$$

$$= \frac{\sqrt{2}}{8.9443} = \boxed{0.1581}$$

so $v = \alpha(A_1-\beta e^{(1)}) = 0.1581\left(\begin{pmatrix}3\\4\end{pmatrix}+5\begin{pmatrix}1\\0\end{pmatrix}\right)$

$$= 0.1581\begin{pmatrix}8\\4\end{pmatrix} = \begin{pmatrix}1.2648\\0.6324\end{pmatrix}$$

so first Unitary factor $U_1 = I - vv^* = \begin{pmatrix}1&0\\0&1\end{pmatrix} - \begin{pmatrix}1.2648\\0.6324\end{pmatrix}(v^H)$

$$= \begin{pmatrix}1&0\\0&1\end{pmatrix} - \begin{pmatrix}1.2648\\0.6324\end{pmatrix}\begin{pmatrix}1.2648 & 0.6324\end{pmatrix}$$

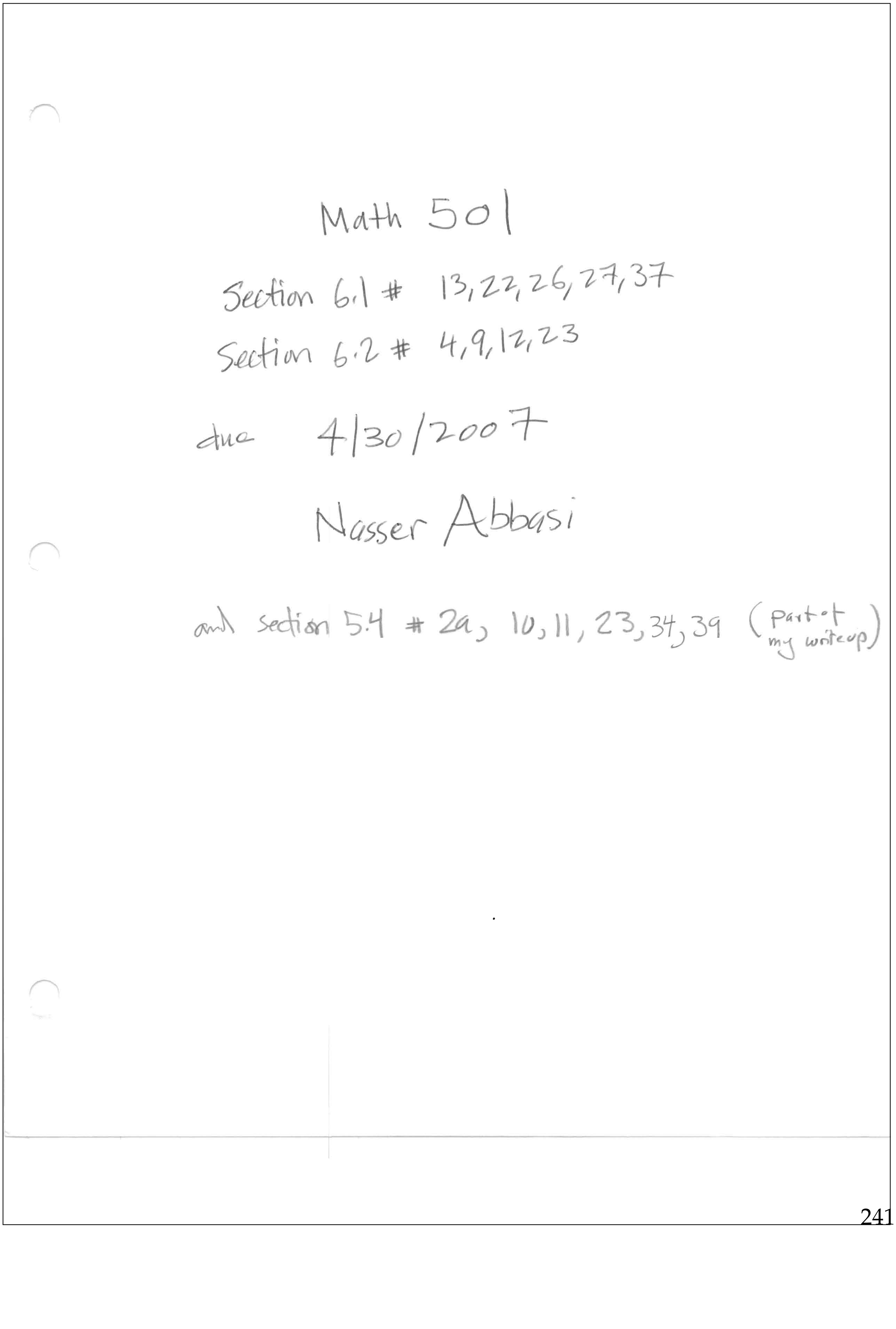
$$U_1 = \begin{pmatrix}-0.6&-0.8\\-0.8&0.6\end{pmatrix}$$

$$U_1A = \begin{pmatrix}-5&-5.2&-6.6\\0&1.4&1.2\end{pmatrix}$$

so $\boxed{Q = U_1^H = \begin{pmatrix}-0.6&-0.8\\-0.8&0.6\end{pmatrix}}$

$$R = U_1A = \begin{pmatrix}-5&-5.2&-6.6\\0&1.4&1.2\end{pmatrix}$$

4.12 HW 11



Math 501

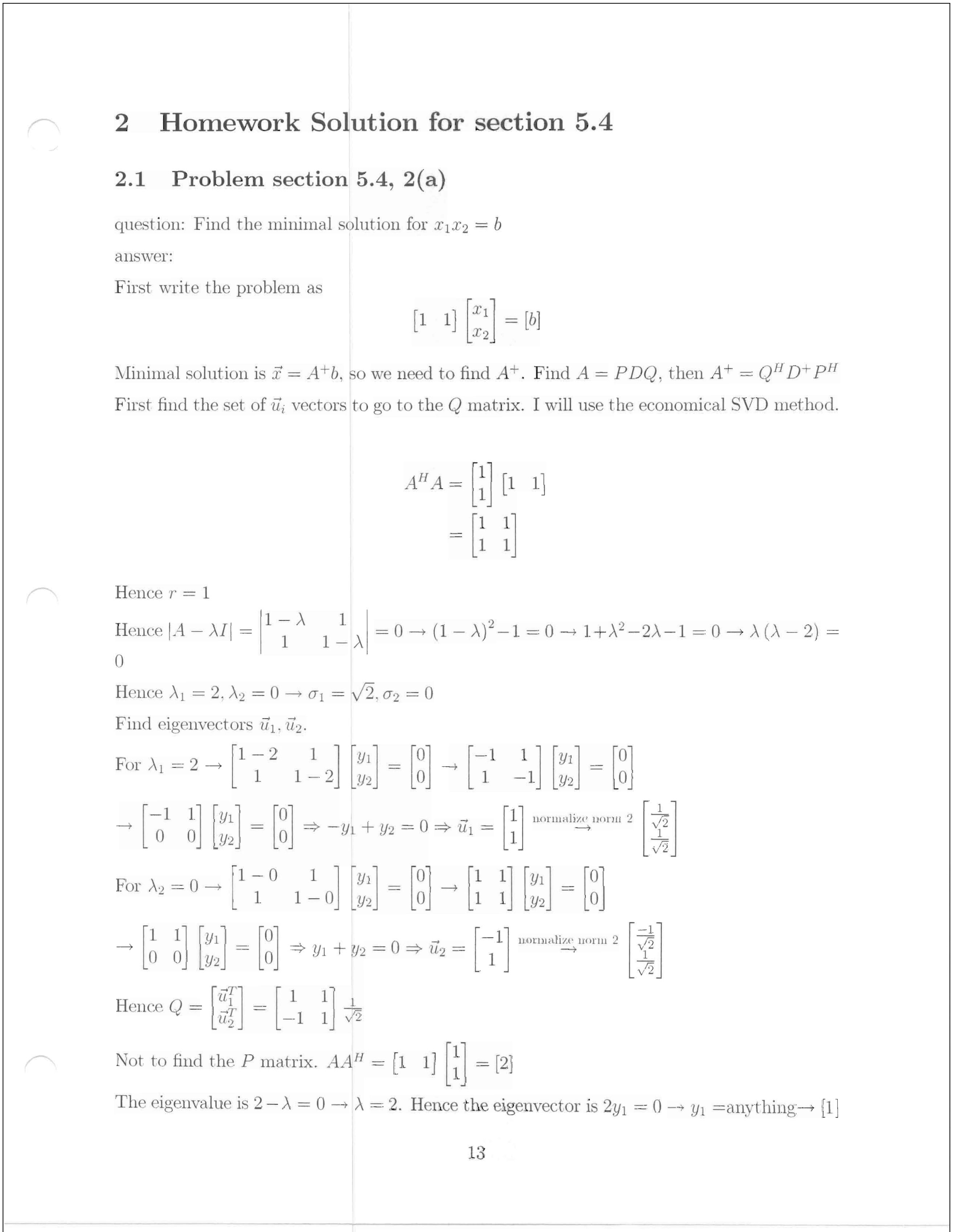
Section 6.1 # 13, 22, 26, 27, 37

Section 6.2 # 4, 9, 12, 23

due 4/30/2007

Nasser Abbasi

and section 5.4 # 2a, 10, 11, 23, 34, 39 (part of my writeup)



2 Homework Solution for section 5.4

2.1 Problem section 5.4, 2(a)

question: Find the minimal solution for $x_1 x_2 = b$

answer:

First write the problem as

$$\begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} = [b]$$

Minimal solution is $\vec{x} = A^{+}b$, so we need to find A^{+}. Find $A = PDQ$, then $A^{+} = Q^{H}D^{+}P^{H}$

First find the set of $\vec{u}_i$ vectors to go to the Q matrix. I will use the economical SVD method.

$$\begin{aligned} A^{H}A &= \begin{bmatrix}1\\ 1\end{bmatrix}\begin{bmatrix}1 & 1\end{bmatrix}\\ &= \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\end{aligned}$$

Hence $r = 1$

Hence $|A - \lambda I| = \begin{vmatrix}1-\lambda & 1\\ 1 & 1-\lambda\end{vmatrix} = 0 \to (1-\lambda)^2 - 1 = 0 \to 1 + \lambda^2 - 2\lambda - 1 = 0 \to \lambda(\lambda - 2) = 0$

Hence $\lambda_1 = 2, \lambda_2 = 0 \to \sigma_1 = \sqrt{2}, \sigma_2 = 0$

Find eigenvectors $\vec{u}_1, \vec{u}_2$.

For $\lambda_1 = 2 \to \begin{bmatrix}1-2 & 1\\ 1 & 1-2\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \to \begin{bmatrix}-1 & 1\\ 1 & -1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$

$\to \begin{bmatrix}-1 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \Rightarrow -y_1 + y_2 = 0 \Rightarrow \vec{u}_1 = \begin{bmatrix}1\\ 1\end{bmatrix} \xrightarrow{\text{normalize norm 2}} \begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$

For $\lambda_2 = 0 \to \begin{bmatrix}1-0 & 1\\ 1 & 1-0\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \to \begin{bmatrix}1 & 1\\ 1 & 1\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix}$

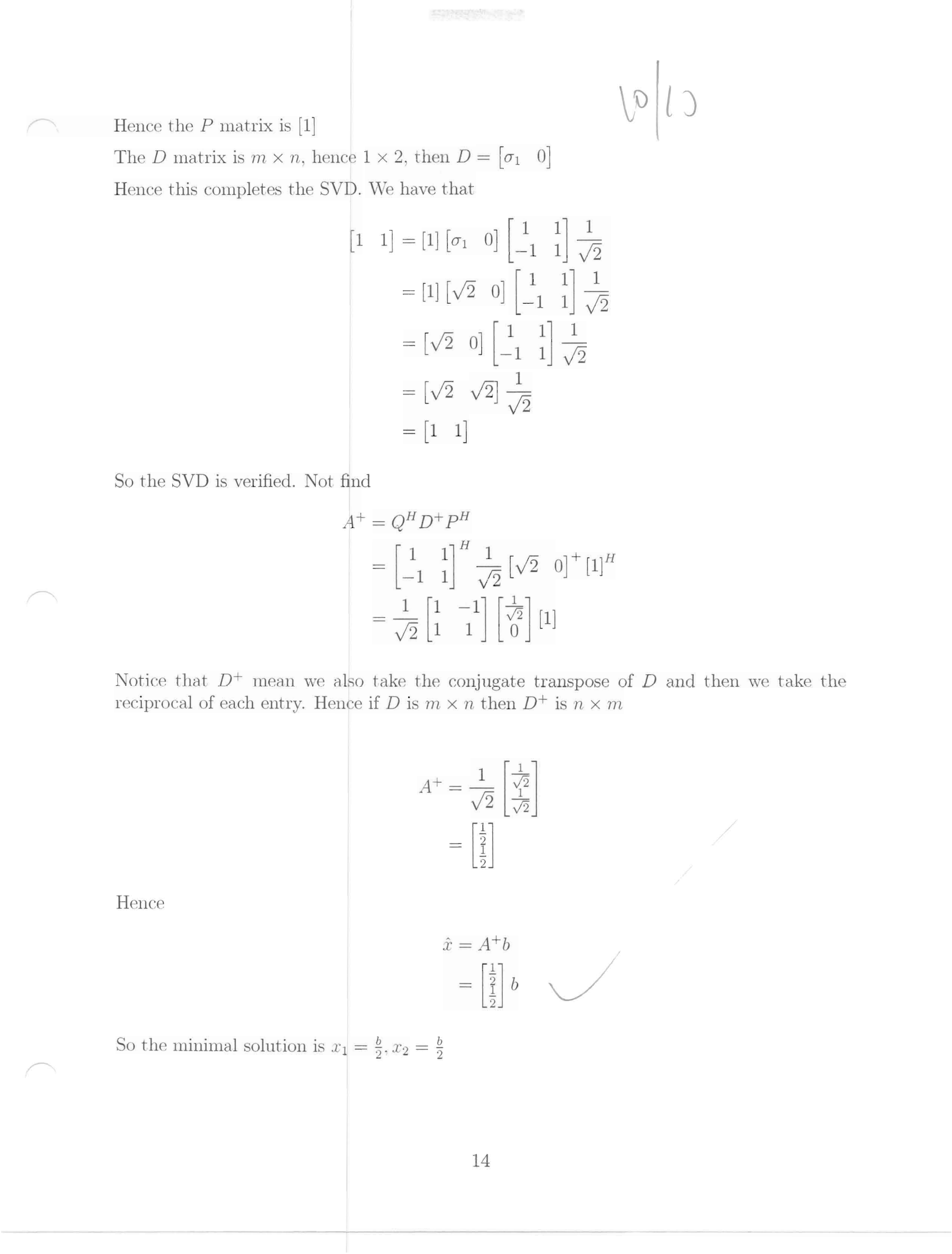
$\to \begin{bmatrix}1 & 1\\ 0 & 0\end{bmatrix}\begin{bmatrix}y_1\\ y_2\end{bmatrix} = \begin{bmatrix}0\\ 0\end{bmatrix} \Rightarrow y_1 + y_2 = 0 \Rightarrow \vec{u}_2 = \begin{bmatrix}-1\\ 1\end{bmatrix} \xrightarrow{\text{normalize norm 2}} \begin{bmatrix}\frac{-1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$

Hence $Q = \begin{bmatrix}\vec{u}_1^T\\ \vec{u}_2^T\end{bmatrix} = \begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}$

Not to find the P matrix. $AA^{H} = \begin{bmatrix}1 & 1\end{bmatrix}\begin{bmatrix}1\\ 1\end{bmatrix} = [2]$

The eigenvalue is $2 - \lambda = 0 \to \lambda = 2$. Hence the eigenvector is $2y_1 = 0 \to y_1 =$anything$\to [1]$

13



10/10

Hence the P matrix is $[1]$

The D matrix is $m \times n$, hence 1×2, then $D = \begin{bmatrix}\sigma_1 & 0\end{bmatrix}$

Hence this completes the SVD. We have that

$$\begin{aligned}\begin{bmatrix}1 & 1\end{bmatrix} &= [1]\begin{bmatrix}\sigma_1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= [1]\begin{bmatrix}\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}\sqrt{2} & \sqrt{2}\end{bmatrix}\frac{1}{\sqrt{2}}\\ &= \begin{bmatrix}1 & 1\end{bmatrix}\end{aligned}$$

So the SVD is verified. Not find

$$\begin{aligned}A^{+} &= Q^{H}D^{+}P^{H}\\ &= \begin{bmatrix}1 & 1\\ -1 & 1\end{bmatrix}^{H}\frac{1}{\sqrt{2}}\begin{bmatrix}\sqrt{2} & 0\end{bmatrix}^{+}[1]^{H}\\ &= \frac{1}{\sqrt{2}}\begin{bmatrix}1 & -1\\ 1 & 1\end{bmatrix}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ 0\end{bmatrix}[1]\end{aligned}$$

Notice that D^{+} mean we also take the conjugate transpose of D and then we take the reciprocal of each entry. Hence if D is $m \times n$ then D^{+} is $n \times m$

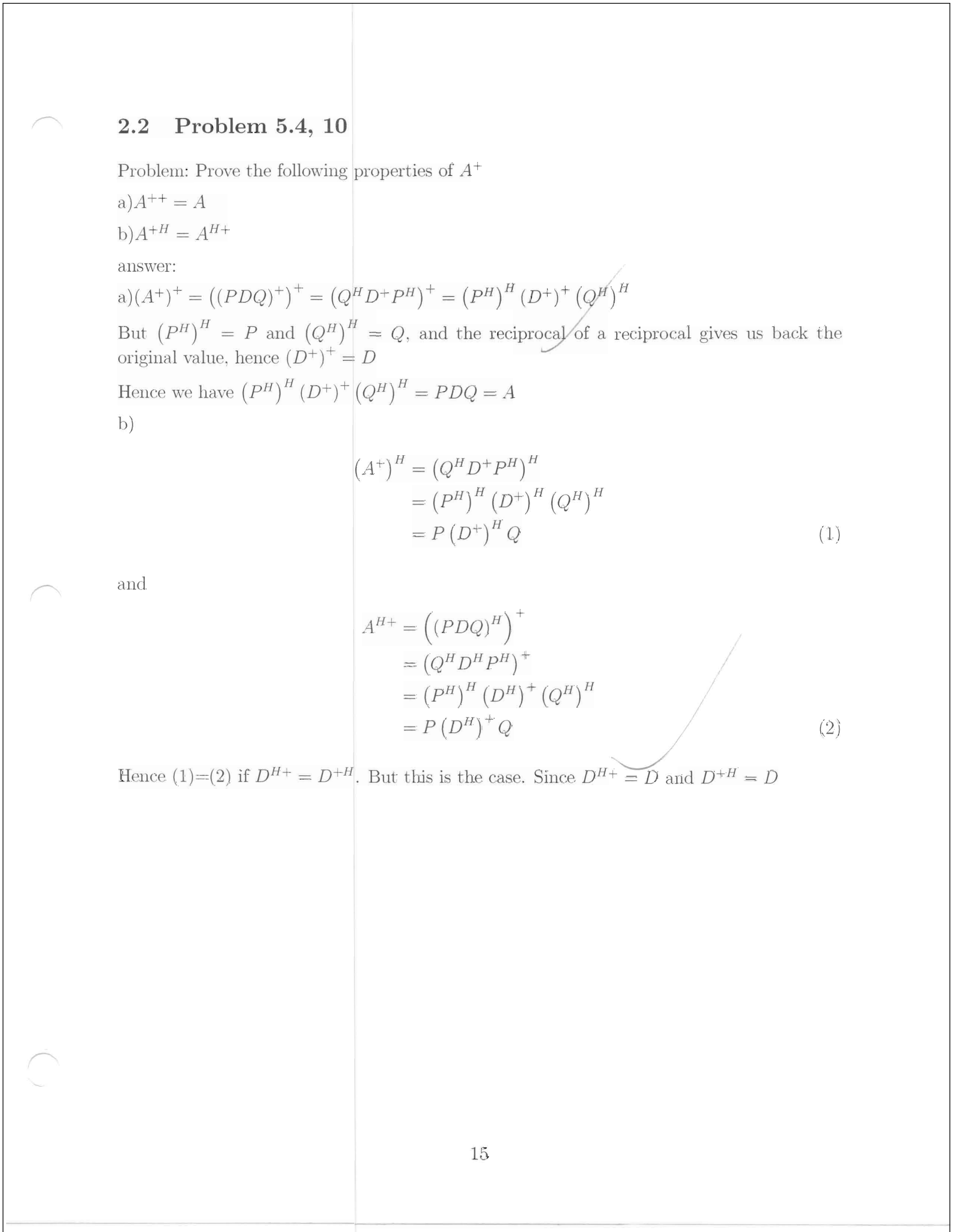
$$\begin{aligned}A^{+} &= \frac{1}{\sqrt{2}}\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}\\ &= \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}\end{aligned}$$

Hence

$$\begin{aligned}\hat{x} &= A^{+}b\\ &= \begin{bmatrix}\frac{1}{2}\\ \frac{1}{2}\end{bmatrix}b\end{aligned}$$

So the minimal solution is $x_1 = \frac{b}{2}, x_2 = \frac{b}{2}$

14



2.2 Problem 5.4, 10

Problem: Prove the following properties of A^+

a)$A^{++} = A$

b)$A^{+H} = A^{H+}$

answer:

a)$(A^+)^+ = ((PDQ)^+)^+ = (Q^H D^+ P^H)^+ = (P^H)^H (D^+)^+ (Q^H)^H$

But $(P^H)^H = P$ and $(Q^H)^H = Q$, and the reciprocal of a reciprocal gives us back the original value, hence $(D^+)^+ = D$

Hence we have $(P^H)^H (D^+)^+ (Q^H)^H = PDQ = A$

b)

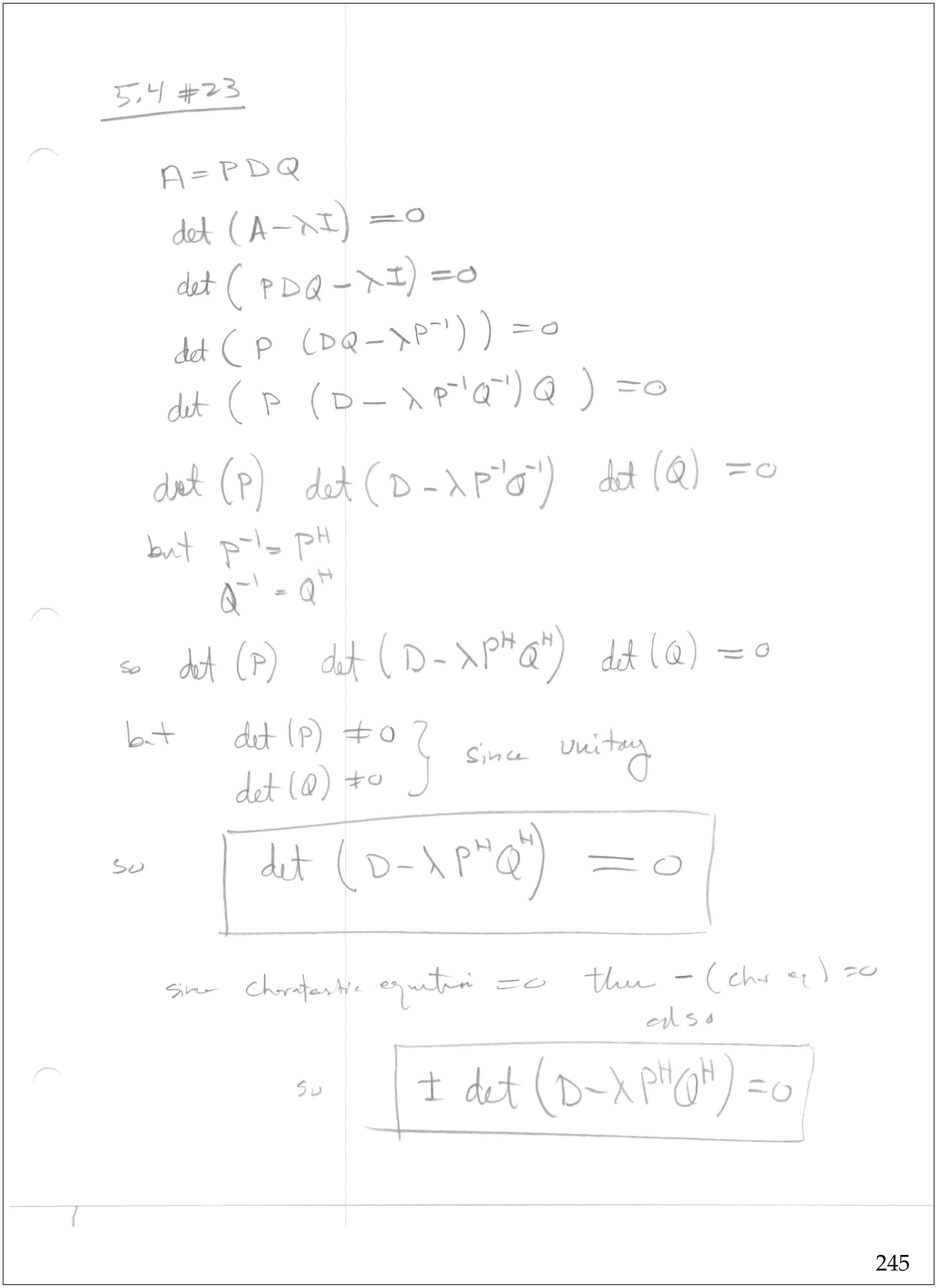
$$\begin{aligned} (A^+)^H &= (Q^H D^+ P^H)^H \\ &= (P^H)^H (D^+)^H (Q^H)^H \\ &= P (D^+)^H Q \qquad (1) \end{aligned}$$

and

$$\begin{aligned} A^{H+} &= \left((PDQ)^H\right)^+ \\ &= (Q^H D^H P^H)^+ \\ &= (P^H)^H (D^H)^+ (Q^H)^H \\ &= P (D^H)^+ Q \qquad (2) \end{aligned}$$

Hence (1)=(2) if $D^{H+} = D^{+H}$. But this is the case. Since $D^{H+} = D$ and $D^{+H} = D$

15



5.4 #23

$A = PDQ$

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

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

$\det(P\ (DQ - \lambda P^{-1})) = 0$

$\det(P\ (D - \lambda P^{-1}Q^{-1})\ Q) = 0$

$\det(P)\ \det(D - \lambda P^{-1}Q^{-1})\ \det(Q) = 0$

but $P^{-1} = P^H$

$Q^{-1} = Q^H$

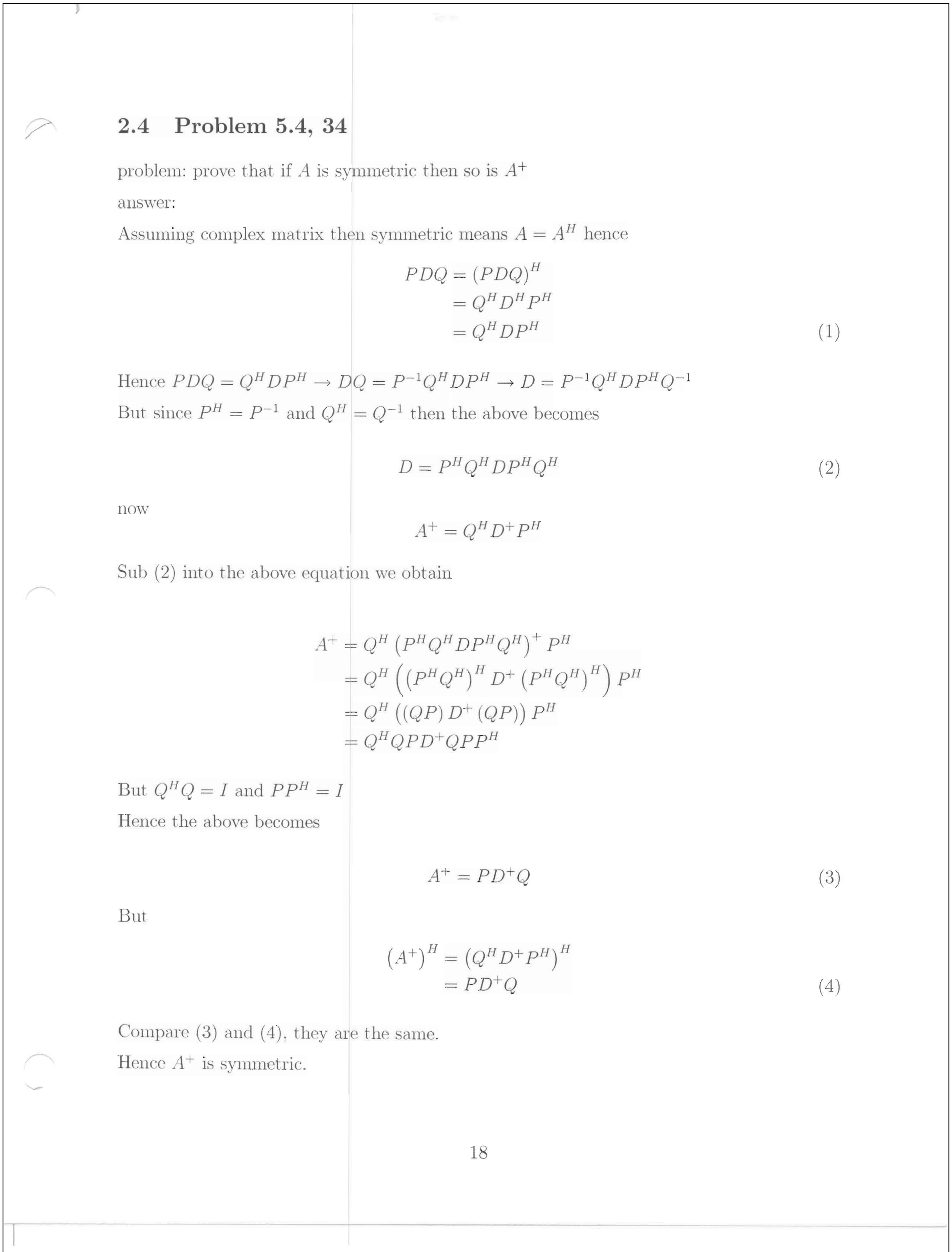
so $\det(P)\ \det(D - \lambda P^H Q^H)\ \det(Q) = 0$

but $\left.\begin{matrix}\det(P) \neq 0 \\ \det(Q) \neq 0\end{matrix}\right\}$ since unitary

so $\boxed{\det(D - \lambda P^H Q^H) = 0}$

since characteristic equation $= 0$ then $-(\text{char eq}) = 0$ also

so $\boxed{\pm \det(D - \lambda P^H Q^H) = 0}$



2.4 Problem 5.4, 34

problem: prove that if A is symmetric then so is A^+

answer:

Assuming complex matrix then symmetric means $A = A^H$ hence

$$\begin{aligned} PDQ &= (PDQ)^H \\ &= Q^H D^H P^H \\ &= Q^H D P^H \end{aligned} \tag{1}$$

Hence $PDQ = Q^H D P^H \rightarrow DQ = P^{-1} Q^H D P^H \rightarrow D = P^{-1} Q^H D P^H Q^{-1}$

But since $P^H = P^{-1}$ and $Q^H = Q^{-1}$ then the above becomes

$$D = P^H Q^H D P^H Q^H \tag{2}$$

now

$$A^+ = Q^H D^+ P^H$$

Sub (2) into the above equation we obtain

$$\begin{aligned} A^+ &= Q^H \left(P^H Q^H D P^H Q^H\right)^+ P^H \\ &= Q^H \left(\left(P^H Q^H\right)^H D^+ \left(P^H Q^H\right)^H\right) P^H \\ &= Q^H \left((QP)\, D^+ \,(QP)\right) P^H \\ &= Q^H Q P D^+ Q P P^H \end{aligned}$$

But $Q^H Q = I$ and $PP^H = I$

Hence the above becomes

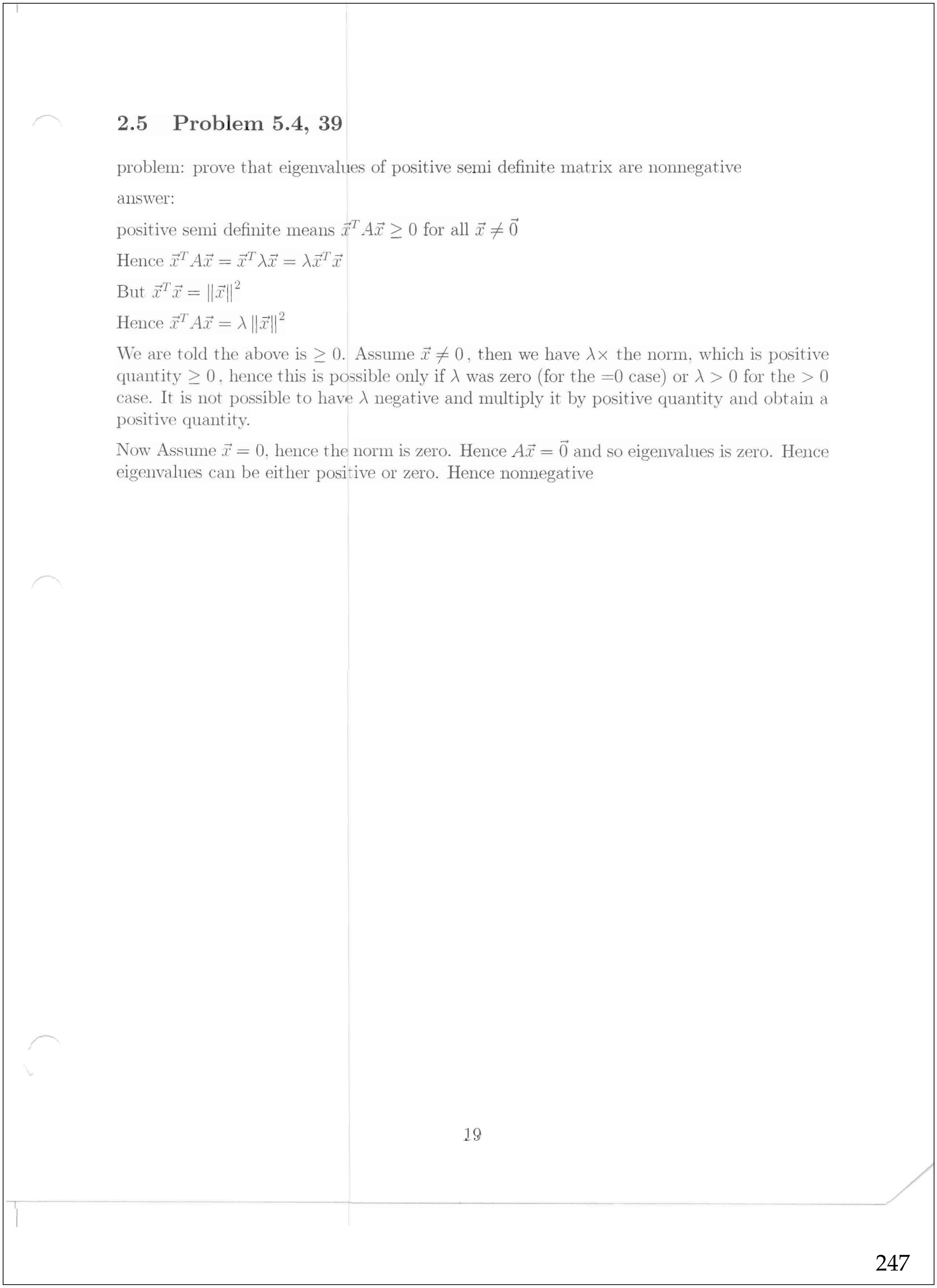
$$A^+ = PD^+ Q \tag{3}$$

But

$$\begin{aligned} \left(A^+\right)^H &= \left(Q^H D^+ P^H\right)^H \\ &= PD^+ Q \end{aligned} \tag{4}$$

Compare (3) and (4), they are the same.

Hence A^+ is symmetric.



2.5 Problem 5.4, 39

problem: prove that eigenvalues of positive semi definite matrix are nonnegative

answer:

positive semi definite means $\vec{x}^T A\vec{x} \geq 0$ for all $\vec{x} \neq \vec{0}$

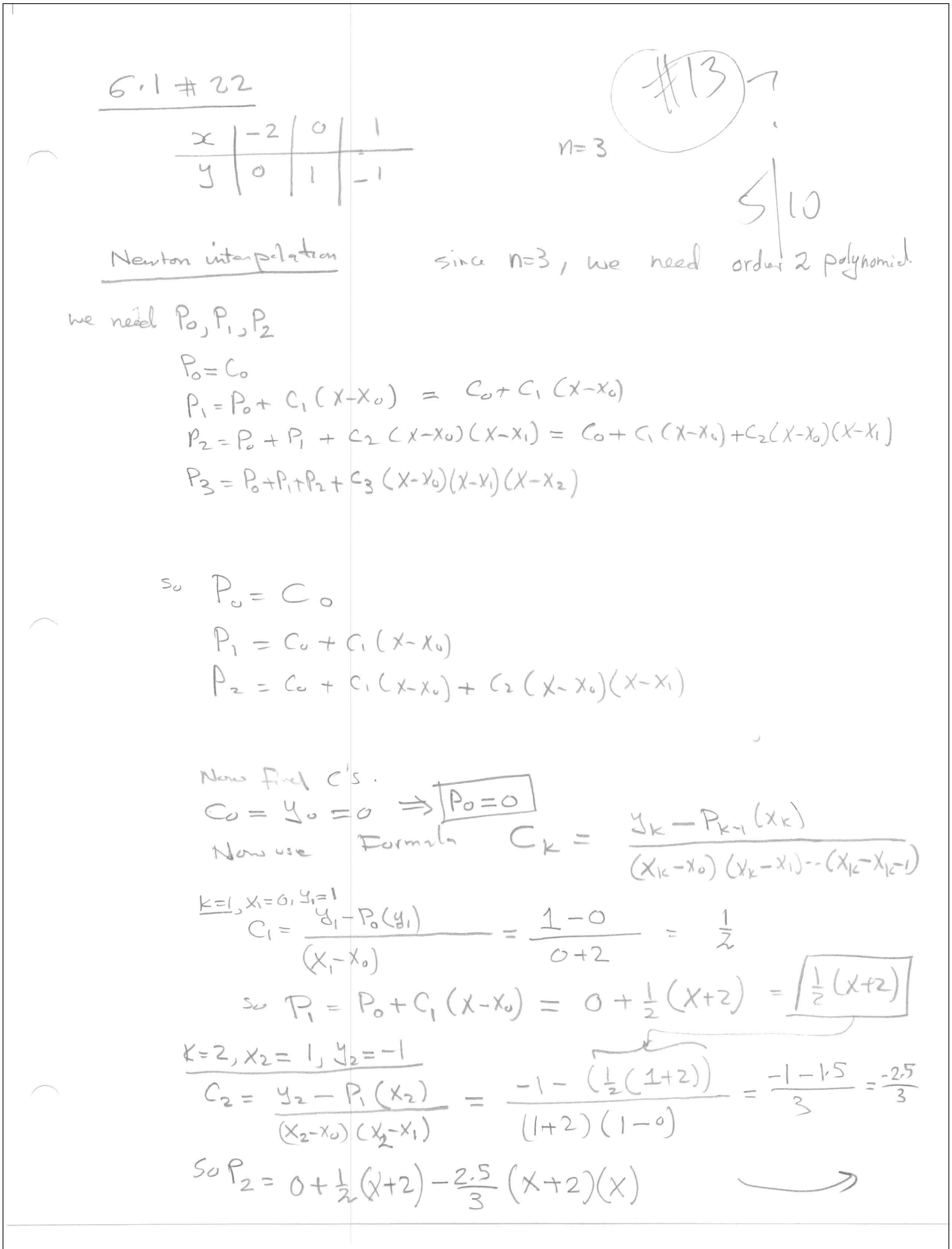
Hence $\vec{x}^T A\vec{x} = \vec{x}^T \lambda\vec{x} = \lambda\vec{x}^T\vec{x}$

But $\vec{x}^T\vec{x} = \|\vec{x}\|^2$

Hence $\vec{x}^T A\vec{x} = \lambda\|\vec{x}\|^2$

We are told the above is ≥ 0. Assume $\vec{x} \neq 0$, then we have $\lambda\times$ the norm, which is positive quantity ≥ 0 , hence this is possible only if λ was zero (for the =0 case) or $\lambda > 0$ for the > 0 case. It is not possible to have λ negative and multiply it by positive quantity and obtain a positive quantity.

Now Assume $\vec{x} = 0$, hence the norm is zero. Hence $A\vec{x} = \vec{0}$ and so eigenvalues is zero. Hence eigenvalues can be either positive or zero. Hence nonnegative



6.1 # 22

x	-2	0	1
y	0	1	-1

$n=3$

#13 ?

5|10

Newton interpolation since $n=3$, we need order 2 polynomial.

we need P_0, P_1, P_2

$$P_0 = C_0$$
$$P_1 = P_0 + C_1(X - X_0) = C_0 + C_1(X - X_0)$$
$$P_2 = P_0 + P_1 + C_2(X - X_0)(X - X_1) = C_0 + C_1(X - X_0) + C_2(X - X_0)(X - X_1)$$
$$P_3 = P_0 + P_1 + P_2 + C_3(X - X_0)(X - X_1)(X - X_2)$$

so
$$P_0 = C_0$$
$$P_1 = C_0 + C_1(X - X_0)$$
$$P_2 = C_0 + C_1(X - X_0) + C_2(X - X_0)(X - X_1)$$

Now find C's.

$C_0 = y_0 = 0 \Rightarrow \boxed{P_0 = 0}$

Now use Formula $C_k = \dfrac{y_k - P_{k-1}(x_k)}{(x_k - x_0)(x_k - x_1)\cdots(x_k - x_{k-1})}$

$k=1$, $x_1=0$, $y_1=1$

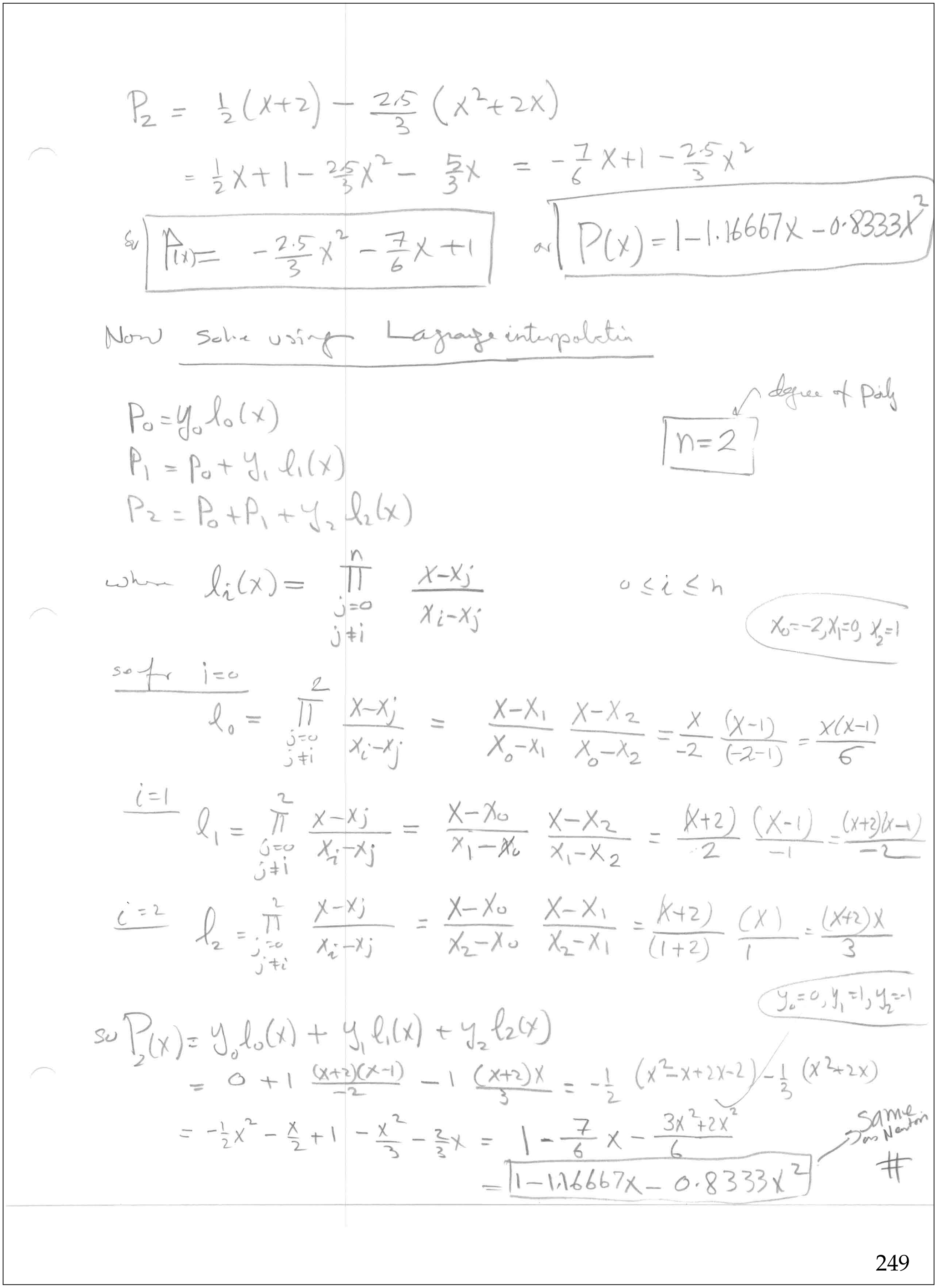
$$C_1 = \frac{y_1 - P_0(y_1)}{(X_1 - X_0)} = \frac{1-0}{0+2} = \frac{1}{2}$$

so $P_1 = P_0 + C_1(X - X_0) = 0 + \frac{1}{2}(X+2) = \boxed{\frac{1}{2}(X+2)}$

$K=2$, $x_2 = 1$, $y_2 = -1$

$$C_2 = \frac{y_2 - P_1(x_2)}{(X_2 - X_0)(X_2 - X_1)} = \frac{-1 - \left(\frac{1}{2}(1+2)\right)}{(1+2)(1-0)} = \frac{-1-1.5}{3} = \frac{-2.5}{3}$$

so $P_2 = 0 + \frac{1}{2}(X+2) - \frac{2.5}{3}(X+2)(X)$ →



$$P_2 = \frac{1}{2}(x+2) - \frac{2.5}{3}(x^2+2x)$$

$$= \frac{1}{2}x + 1 - \frac{2.5}{3}x^2 - \frac{5}{3}x = -\frac{7}{6}x + 1 - \frac{2.5}{3}x^2$$

so $\boxed{P(x) = -\frac{2.5}{3}x^2 - \frac{7}{6}x + 1}$ or $\boxed{P(x) = 1 - 1.16667x - 0.8333x^2}$

Now solve using Lagrange interpolation

$P_0 = y_0 l_0(x)$

$P_1 = P_0 + y_1 l_1(x)$

$P_2 = P_0 + P_1 + y_2 l_2(x)$

$\boxed{n=2}$ degree of Poly

where $l_i(x) = \prod_{\substack{j=0\\ j\neq i}}^{n} \frac{x-x_j}{x_i-x_j}$ $\quad 0 \leq i \leq n$

$x_0=-2, x_1=0, x_2=1$

so for $i=0$

$$l_0 = \prod_{\substack{j=0\\ j\neq i}}^{2} \frac{x-x_j}{x_i-x_j} = \frac{x-x_1}{x_0-x_1}\frac{x-x_2}{x_0-x_2} = \frac{x}{-2}\frac{(x-1)}{(-2-1)} = \frac{x(x-1)}{6}$$

$i=1$

$$l_1 = \prod_{\substack{j=0\\ j\neq i}}^{2} \frac{x-x_j}{x_i-x_j} = \frac{x-x_0}{x_1-x_0}\frac{x-x_2}{x_1-x_2} = \frac{(x+2)}{2}\frac{(x-1)}{-1} = \frac{(x+2)(x-1)}{-2}$$

$i=2$

$$l_2 = \prod_{\substack{j=0\\ j\neq i}}^{2} \frac{x-x_j}{x_i-x_j} = \frac{x-x_0}{x_2-x_0}\frac{x-x_1}{x_2-x_1} = \frac{(x+2)}{(1+2)}\frac{(x)}{1} = \frac{(x+2)x}{3}$$

$y_0=0, y_1=1, y_2=-1$

$$\text{so } P_2(x) = y_0 l_0(x) + y_1 l_1(x) + y_2 l_2(x)$$

$$= 0 + 1\,\frac{(x+2)(x-1)}{-2} - 1\,\frac{(x+2)x}{3} = -\frac{1}{2}(x^2 - x + 2x - 2) - \frac{1}{3}(x^2+2x)$$

$$= -\frac{1}{2}x^2 - \frac{x}{2} + 1 - \frac{x^2}{3} - \frac{2}{3}x = 1 - \frac{7}{6}x - \frac{3x^2+2x^2}{6}$$

$$= \boxed{1 - 1.16667x - 0.8333x^2}$$

same as Newton

#

6.1 #26

$x - 9^{-x} = 0 \Rightarrow$

x	0	0.5	1
y	$f(x)$	$f(x)$	$f(x)$

where $f(x) = x - 9^{-x}$

$\Rightarrow$

x	0	0.5	1
y	-1	0.166667	0.888889

Use Newton

$$P = C_0 + C_1(x - x_0) + C_2(x - x_0)(x - x_1)$$

$$C_0 = y_0 = -1$$

$$C_1 = \frac{y_1 - P_0(x_1)}{x_1 - x_0} = \frac{0.166667 + 1}{0.5} = 2.3333$$

$\Rightarrow P_1 = C_1(x - x_0)$

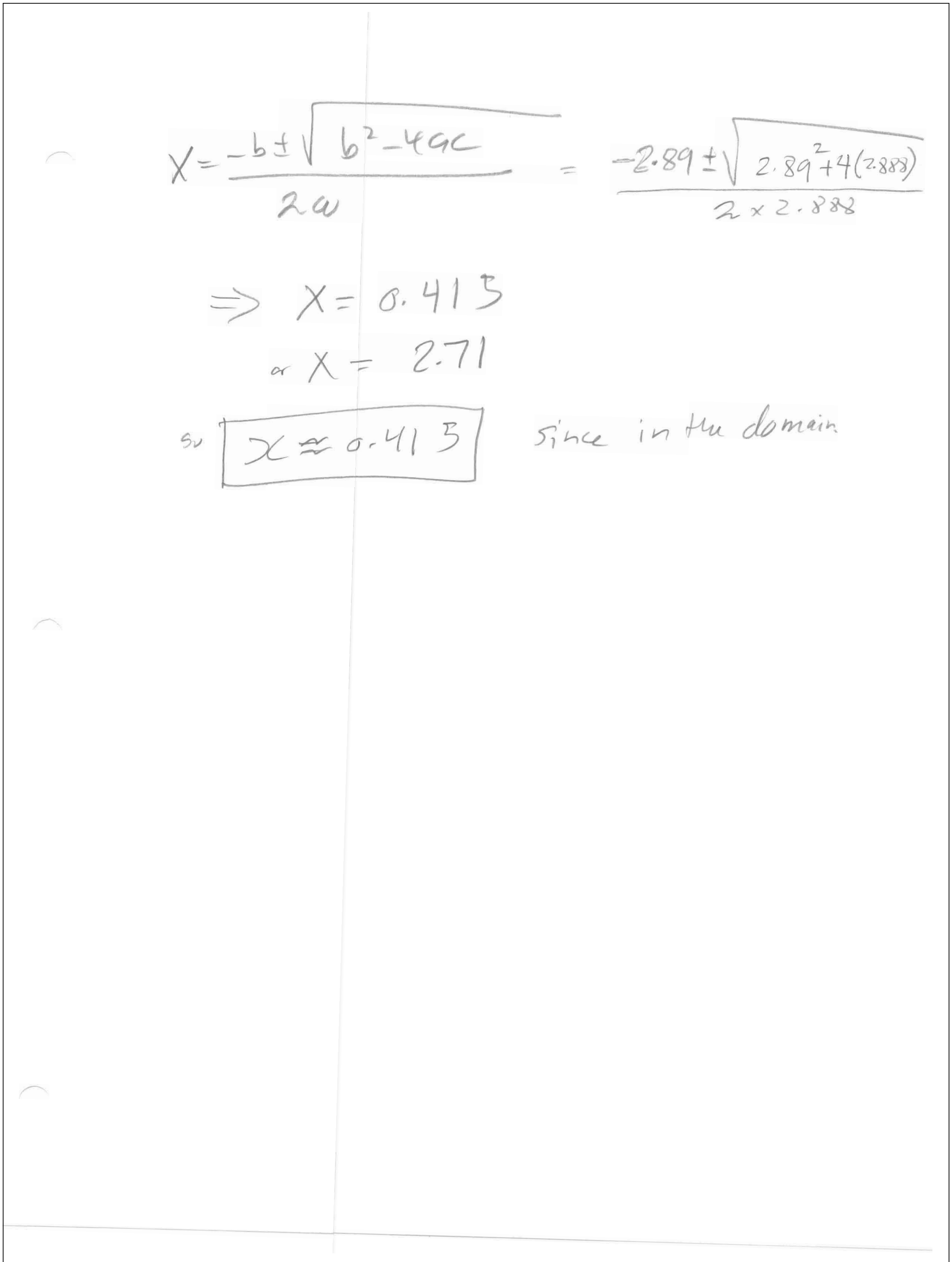
$$C_2 = \frac{y_2 - P_1(x_2)}{(x_2 - x_0)(x_2 - x_1)} = \frac{0.88889 - 2.3333(1)}{(1-0)(1-0.5)} = 2.8888$$

$$\Rightarrow P_2(x) = 2.8888(x - 0)(x - 0.5) = 2.8888(x)(x - 0.5)$$

so $P(x) = -1 + 2.3333(x) + 2.88888(x)(x - 0.5)$

$$\boxed{P(x) = -1 + 2.89x + 2.888x^2}$$

Note (using exact rational values is more accurate numerically, but by hand faster to use calculator and use decimal points)



$$X = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-2.89 \pm \sqrt{2.89^2 + 4(2.888)}}{2 \times 2.888}$$

$$\Rightarrow X = 0.413$$

$$\text{or } X = 2.71$$

so $\boxed{x \approx 0.413}$ since in the domain

6.1 # 27

if we interpolate $f(x)=e^{x-1}$ with polynomial p of degree 12 using 13 nodes in $[-1,1]$ what is a good upper bound for $|f(x)-p(x)|$ on $[-1,1]$?

using theorem 2, page 315 on interpolation error:

$$f(x)-p(x)=\frac{1}{(n+1)!}f^{(n+1)}(\xi_x)\prod_{i=0}^{n}(x-x_i)$$

$$\prod_{i=0}^{12}|x-x_i| \le 2^{12}$$

$$f(x)=e^{x-1}$$
$$f^{(1)}(x)=e^{x-1}$$
$$f^{(2)}(x)=e^{x-1} \quad \text{so} \quad f^{(13)}(x)=e^{x-1}$$

$$\text{so } |f^{(13)}(\xi_x)| \le e^1$$

$$\text{so } |f(x)-p(x)| \le \frac{(e^1)(2^{12})}{(13)!} \le \boxed{1.78803\times10^{-6}}$$

6.1

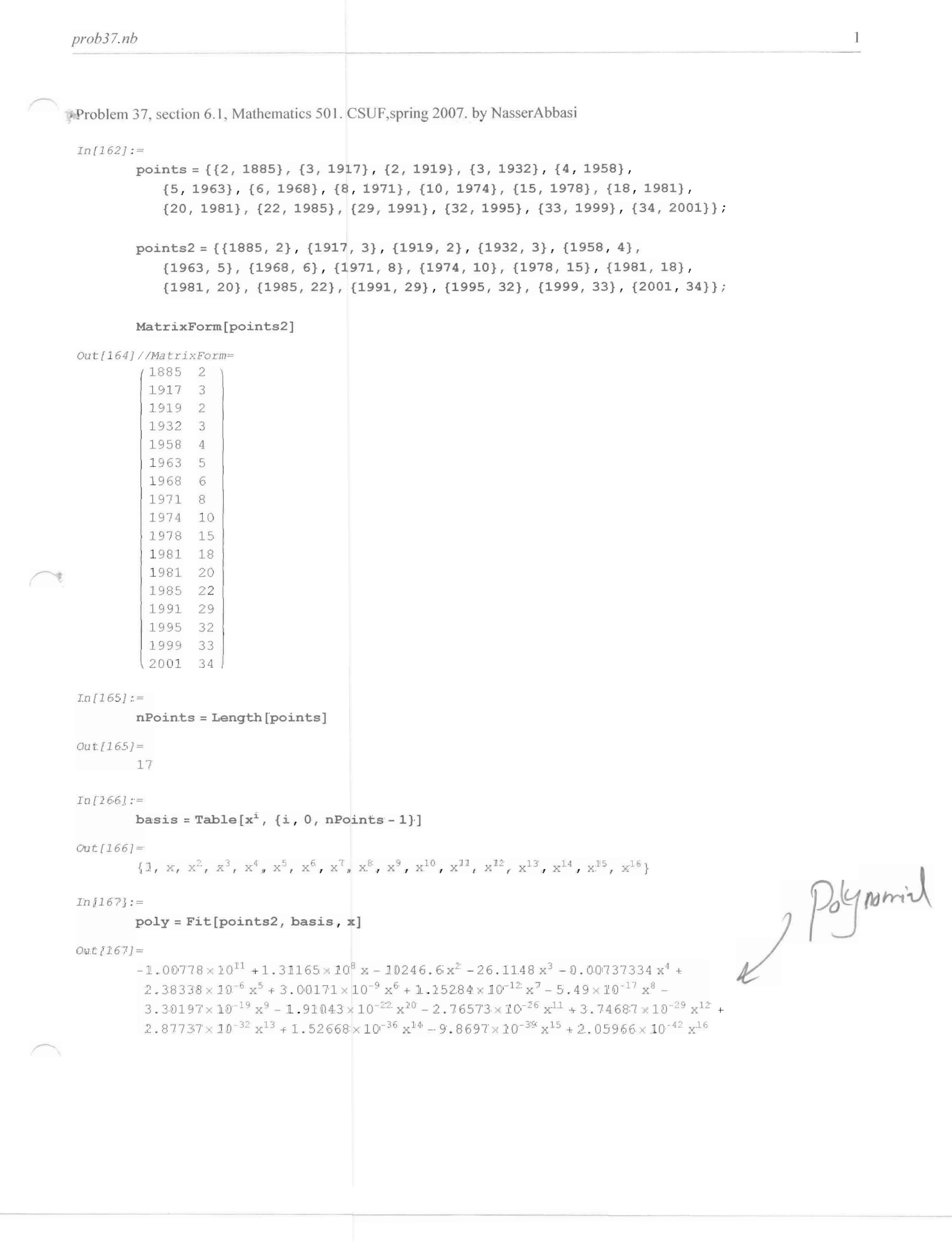
37

cost	2	3	2	3	4	5	6	8	10
year	1885	1917	1919	1932	1958	1963	1968	1971	1974

15	18	20	22	25	29	32	33	34.
1978	1981	1981	1985	1988	1991	1995	1999	2001

find Newton polynomial. when will it cost $1 to mail a letter? when will it cost $10 ?

→ solution back.



Problem 37, section 6.1, Mathematics 501. CSUF,spring 2007. by NasserAbbasi

```
In[162]:=
        points = {{2, 1885}, {3, 1917}, {2, 1919}, {3, 1932}, {4, 1958},
           {5, 1963}, {6, 1968}, {8, 1971}, {10, 1974}, {15, 1978}, {18, 1981},
           {20, 1981}, {22, 1985}, {29, 1991}, {32, 1995}, {33, 1999}, {34, 2001}};

        points2 = {{1885, 2}, {1917, 3}, {1919, 2}, {1932, 3}, {1958, 4},
           {1963, 5}, {1968, 6}, {1971, 8}, {1974, 10}, {1978, 15}, {1981, 18},
           {1981, 20}, {1985, 22}, {1991, 29}, {1995, 32}, {1999, 33}, {2001, 34}};

        MatrixForm[points2]

Out[164]//MatrixForm=
        1885  2
        1917  3
        1919  2
        1932  3
        1958  4
        1963  5
        1968  6
        1971  8
        1974  10
        1978  15
        1981  18
        1981  20
        1985  22
        1991  29
        1995  32
        1999  33
        2001  34

In[165]:=
        nPoints = Length[points]

Out[165]=
        17

In[166]:=
        basis = Table[x^i, {i, 0, nPoints - 1}]
```

Out[166]=

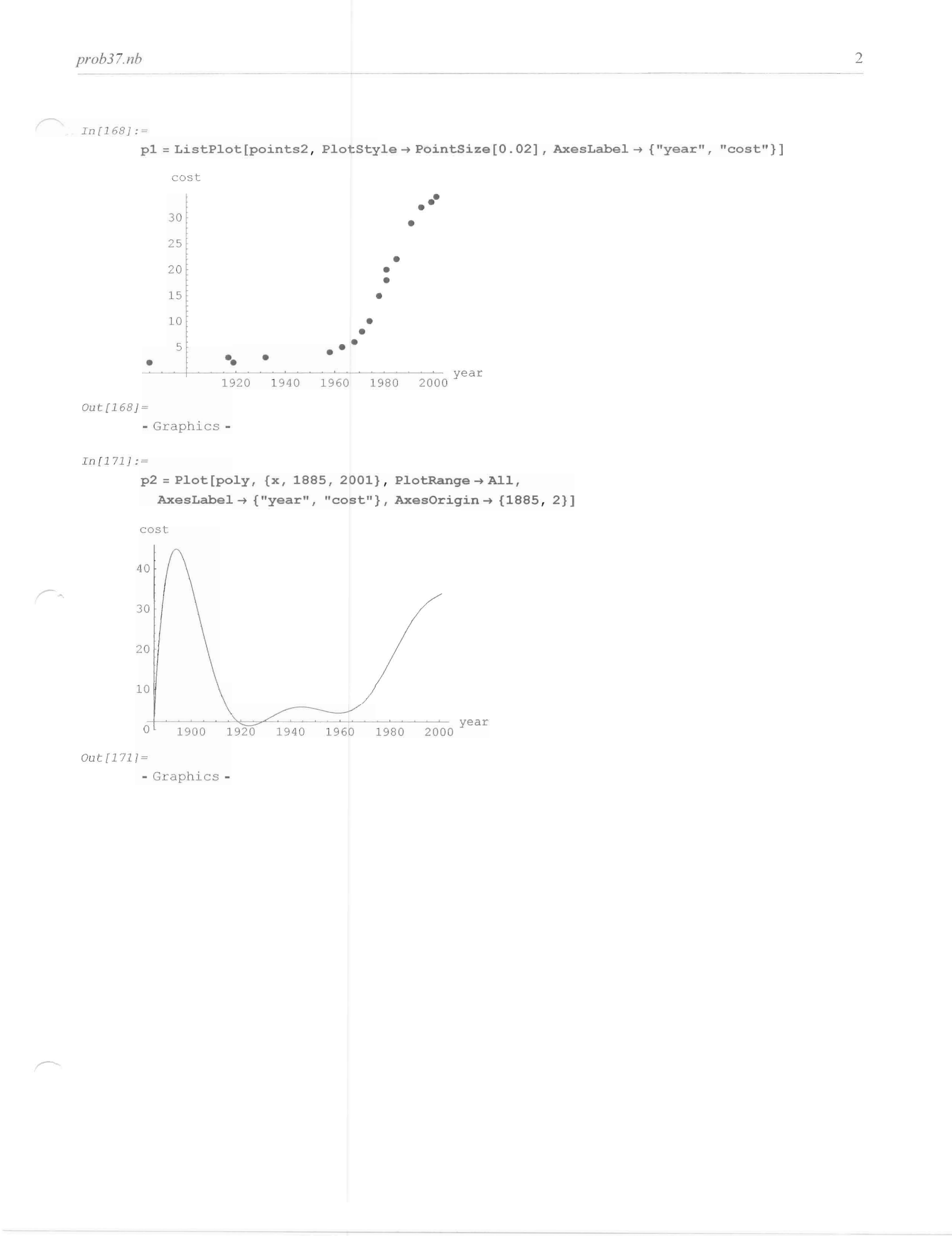
$$\{1, x, x^2, x^3, x^4, x^5, x^6, x^7, x^8, x^9, x^{10}, x^{11}, x^{12}, x^{13}, x^{14}, x^{15}, x^{16}\}$$

```
In[167]:=
        poly = Fit[points2, basis, x]
```

Polynomial

Out[167]=

$$-1.00778\times10^{11} + 1.31165\times10^{8}\,x - 10246.6\,x^2 - 26.1148\,x^3 - 0.00737334\,x^4 + 2.38338\times10^{-6}\,x^5 + 3.00171\times10^{-9}\,x^6 + 1.15284\times10^{-12}\,x^7 - 5.49\times10^{-17}\,x^8 - 3.30197\times10^{-19}\,x^9 - 1.91043\times10^{-22}\,x^{10} - 2.76573\times10^{-26}\,x^{11} + 3.74687\times10^{-29}\,x^{12} + 2.87737\times10^{-32}\,x^{13} + 1.52668\times10^{-36}\,x^{14} - 9.8697\times10^{-39}\,x^{15} + 2.05966\times10^{-42}\,x^{16}$$



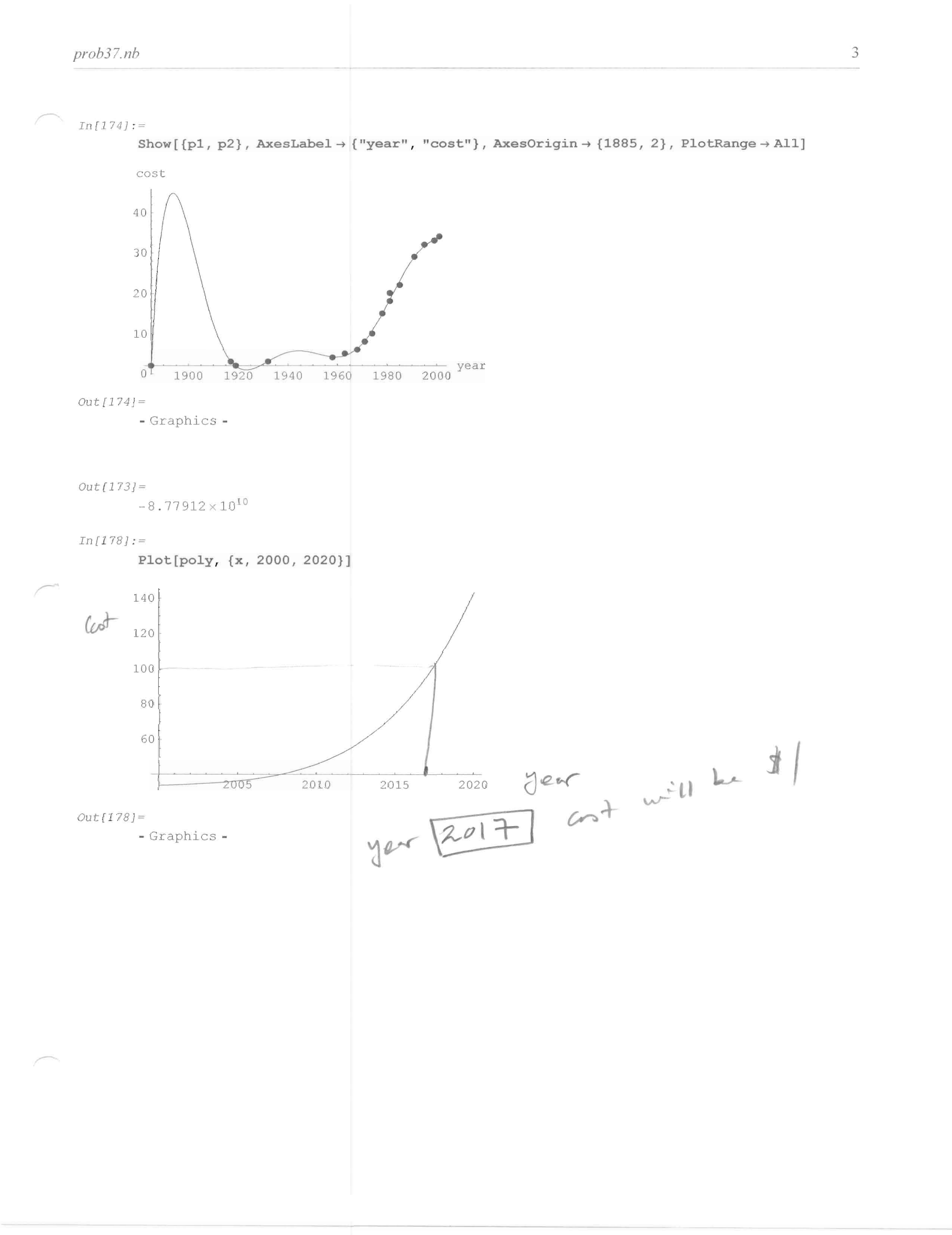
prob37.nb 2

```
In[168]:=
        p1 = ListPlot[points2, PlotStyle → PointSize[0.02], AxesLabel → {"year", "cost"}]
```

```
Out[168]=
        - Graphics -
```

```
In[171]:=
        p2 = Plot[poly, {x, 1885, 2001}, PlotRange → All,
          AxesLabel → {"year", "cost"}, AxesOrigin → {1885, 2}]
```

```
Out[171]=
        - Graphics -
```



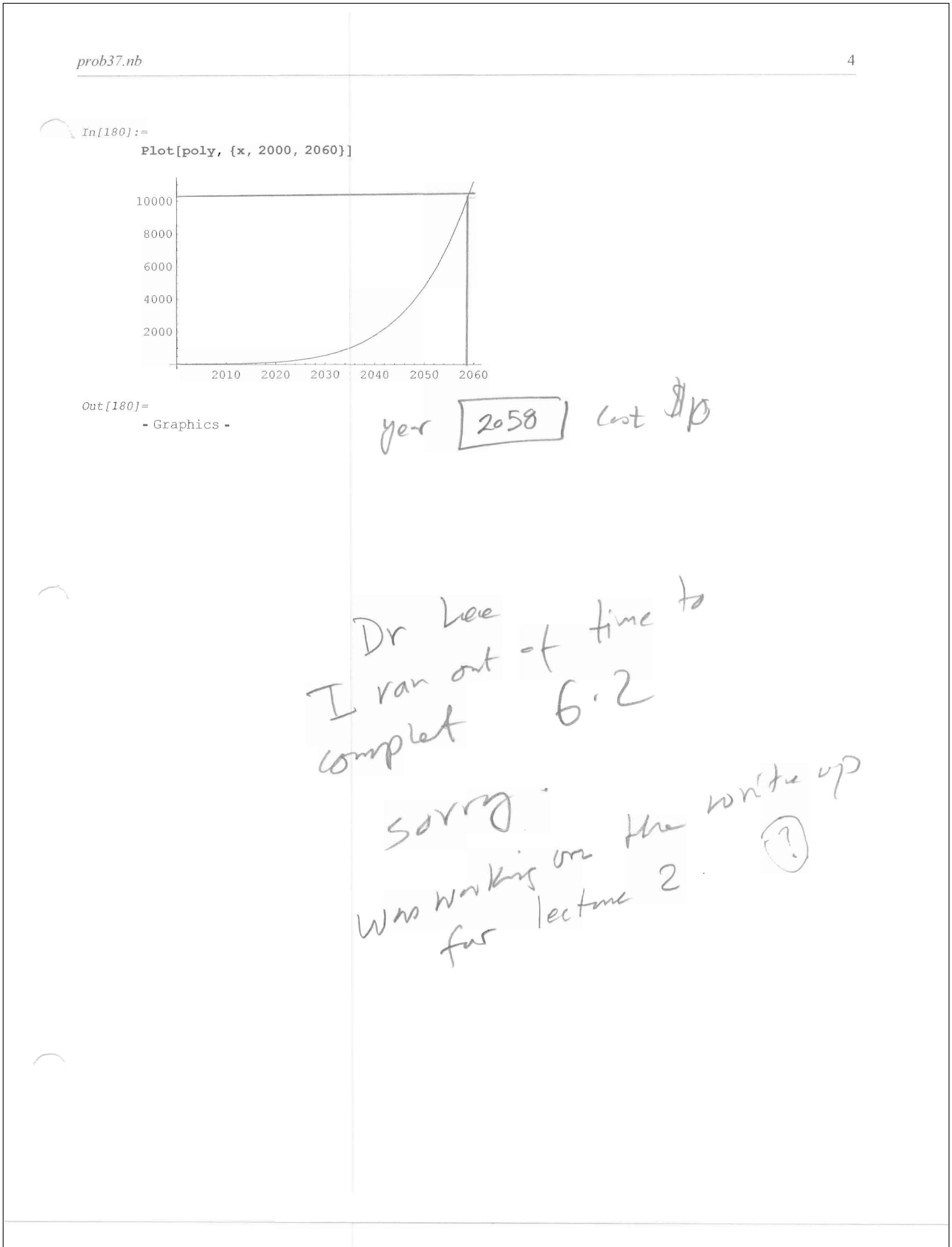
prob37.nb 3

```
In[174]:=
        Show[{p1, p2}, AxesLabel → {"year", "cost"}, AxesOrigin → {1885, 2}, PlotRange → All]
```

```
Out[174]=
        - Graphics -


Out[173]=
        -8.77912×10^10

In[178]:=
        Plot[poly, {x, 2000, 2020}]
```

```
Out[178]=
        - Graphics -
```

year $\boxed{2017}$ cost will be $1


10000
8000
6000
4000
2000
2010
2020
2030
2040
2050
2060

prob37.nb 4

```
In[180]:=
        Plot[poly, {x, 2000, 2060}]
```

```
Out[180]=
        - Graphics -
```

year $\boxed{2058}$ cost $10

Dr Lee
I ran out of time to
complet 6.2

sorry.
was working on the write up
for lecture 2 ?

4.13 HW 12

Local contents

4.13.1 Section 7.1, Problem 6

Problem: Derive the following 2 formulas for approximation of derivatives and show they are both $O(h^4)$ by evaluating their error terms

$$\begin{aligned} f'(x) &= \frac{1}{12h}\left[-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)\right] \\ f''(x) &= \frac{1}{12h^2}\left[-f(x+2h)+16f(x+h)-30f(x)+16f(x-h)-f(x-2h)\right] \end{aligned}$$

Solution:

I could obtain the above results directly from applying Richardson interpolation formulas (which is a short approach), but I assumed the question wanted us to derive these from first principles. I first show how to do one using Richardson, then solve both from first principles.

To obtain the approximation for $f'(x)$ using Richardson, we do the following:

$$\begin{aligned} \varphi(h) &= \frac{1}{2h}\left[f(x+h)-f(x-h)\right] \\ L &= \varphi(h)+a_2h^2+a_4h^4+\cdots \end{aligned} \tag{1C}$$

Replacing h by $2h$

$$L = \varphi(2h)+a_2 4h^2+a_4 16h^4+\cdots \tag{2C}$$

Multiplying (1C) by 4 and subtract (2C) from result

$$\begin{aligned} 3L &= \left(4\varphi(h)+4a_2h^2+4a_4h^4+\cdots\right)-\left(\varphi(2h)+a_2 4h^2+a_4 16h^4+\cdots\right) \\ &= 4\varphi(h)-\varphi(2h)-12a_4h^4-\cdots \end{aligned}$$

Hence

$$\begin{aligned} L &= \frac{1}{3}\left(\frac{2}{h}\left[f(x+h)-f(x-h)\right]-\frac{1}{4h}\left[f(x+2h)-f(x-2h)\right]-12a_4h^4-\cdots\right) \\ &= \frac{2}{3h}\left[f(x+h)-f(x-h)\right]-\frac{1}{12h}\left[f(x+2h)-f(x-2h)\right]-4a_4h^4-\cdots \\ &= \frac{1}{12h}\left(8\left[f(x+h)-f(x-h)\right]-\left[f(x+2h)-f(x-2h)\right]\right)-4a_4h^4-\cdots \\ &= \frac{1}{12h}\left[-f(x+2h)+8f(x+h)-8f(x-h)+f(x-2h)\right]-4a_4h^4-\cdots \end{aligned}$$

Which is the same result obtained earlier using the long approach. We also see that the error term is $O\left(h^4\right)$

Now, solve it again, but using direct usage of Taylor (which I assume what the book wanted us to do)

From Taylor expansion, we write, by expanding around $x+h$ and $x-h$

$$\begin{aligned}
f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) + \frac{h^5}{5!}f^{(5)}(\xi_1) \\
f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) - \frac{h^5}{5!}f^{(5)}(\xi_2)
\end{aligned}$$

Subtract the second from the first equation

$$f(x+h) - f(x-h) = 2hf'(x) + \frac{h^3}{3}f^{(3)}(x) + \frac{h^5}{60}\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right]$$

Solve for $f'(x)$ we obtain

$$f'(x) = \frac{1}{2h}\left[f(x+h) - f(x-h)\right] - \frac{1}{6}h^2 f^{(3)}(x) - \frac{1}{120}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right] \tag{1}$$

Now we do the same again, but by expanding around $x+2h$ and $x-2h$

$$\begin{aligned}
f(x+2h) &= f(x) + 2hf'(x) + \frac{(2h)^2}{2}f''(x) + \frac{(2h)^3}{3!}f^{(3)}(x) + \frac{(2h)^4}{4!}f^{(4)}(x) + \frac{(2h)^5}{5!}f^{(5)}(\xi_1) \\
f(x-2h) &= f(x) - 2hf'(x) + \frac{(2h)^2}{2}f''(x) - \frac{(2h)^3}{3!}f^{(3)}(x) + \frac{(2h)^4}{4!}f^{(4)}(x) - \frac{(2h)^5}{5!}f^{(5)}(\xi_2)
\end{aligned}$$

Subtract the second from the first equation

$$\begin{aligned}
f(x+2h) - f(x-2h) &= 4hf'(x) + 2\frac{(2h)^3}{3!}f^{(3)}(x) + \frac{(2h)^5}{5!}\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right] \\
&= 4hf'(x) + \frac{8}{3}h^3 f^{(3)}(x) + \frac{4}{15}h^5\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right]
\end{aligned}$$

Solve for $f'(x)$ we obtain

$$f\,'(x)=\frac{1}{4h}\left[f\left(x+2h\right)-f\left(x-2h\right)\right]-\frac{4}{6}h^{2}f^{(3)}\left(x\right)-\frac{1}{15}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right] \tag{2}$$

We want to eliminate $f^{(3)}(x)$ from the above. So we multiply eq(1) by 4 and subtract eq(2) from the result. So equation (1) becomes

$$\begin{aligned}4f\,'(x)&=4\left(\frac{1}{2h}\left[f\left(x+h\right)-f\left(x-h\right)\right]-\frac{1}{6}h^{2}f^{(3)}(x)-\frac{1}{120}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]\right)\\&=\frac{2}{h}\left[f\left(x+h\right)-f\left(x-h\right)\right]-\frac{4}{6}h^{2}f^{(3)}(x)-\frac{1}{30}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right] \qquad (3)\end{aligned}$$

Now subtract (2) from (3) we obtain

$$\begin{aligned}3f'(x)=\frac{2}{h}&\left[f\left(x+h\right)-f\left(x-h\right)\right]-\frac{4}{6}h^{2}f^{(3)}\left(x\right)-\frac{1}{30}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]-\\&\left(\frac{1}{4h}\left[f\left(x+2h\right)-f\left(x-2h\right)\right]-\frac{4}{6}h^{2}f^{(3)}(x)-\frac{1}{15}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]\right)\end{aligned}$$

Hence

$$\begin{aligned}3f'(x)&=\frac{2}{h}\left[f\left(x+h\right)-f\left(x-h\right)\right]-\frac{1}{30}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]-\\&\quad\frac{1}{4h}\left[f\left(x+2h\right)-f\left(x-2h\right)\right]+\frac{1}{15}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]\end{aligned}$$

$$\begin{aligned}f'(x)&=\frac{2}{3h}\left[f\left(x+h\right)-f\left(x-h\right)\right]-\frac{1}{12h}\left[f\left(x+2h\right)-f\left(x-2h\right)\right]+\frac{1}{90}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]\\&=\frac{1}{12h}\left[8f\left(x+h\right)-8f\left(x-h\right)-f\left(x+2h\right)-f\left(x-2h\right)\right]+\frac{1}{90}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]\end{aligned}$$

Rearrange terms to make it look as in the textbook

$$f'(x)=\frac{1}{12h}\left[-f\left(x+2h\right)+8f\left(x+h\right)-8f\left(x-h\right)-f\left(x-2h\right)\right]+\frac{1}{90}h^{4}\left[f^{(5)}\left(\xi\right)\right] \tag{4}$$

Where we replaced $\frac{1}{90}h^{4}\left[f^{(5)}\left(\xi_{1}\right)+f^{(5)}\left(\xi_{2}\right)\right]$ by $\frac{1}{45}h^{4}\left[\frac{f^{(5)}(\xi_{1})+f^{(5)}(\xi_{2})}{2}\right]=\frac{1}{90}h^{4}\left[f^{(5)}\left(\xi\right)\right]$ with $f^{(5)}\left(\xi\right)$ being the mean value of $\frac{f^{(5)}(\xi_{1})+f^{(5)}(\xi_{2})}{2}$

Hence from equation (4) we see that the error is $O\left(h^4\right)$ as required to show.

Hence

$$\boxed{f'(x) \approx \frac{1}{12h}\left[-f(x+2h)+8f(x+h)-8f(x-h)-f(x-2h)\right]}$$

Now we need to show the formula for $f''(x)$. We do the same as above, but instead of subtracting equations, we add them. We start from the top to show these again step by step.

From Taylor expansion, we write, by expanding around $x+h$ and $x-h$

$$\begin{aligned} f(x+h) &= f(x)+hf'(x)+\frac{h^2}{2}f''(x)+\frac{h^3}{3!}f^{(3)}(x)+\frac{h^4}{4!}f^{(4)}(x)+\frac{h^5}{5!}f^{(5)}(x)+\frac{h^6}{6!}f^{(6)}(\xi_1) \\ f(x-h) &= f(x)-hf'(x)+\frac{h^2}{2}f''(x)-\frac{h^3}{3!}f^{(3)}(x)+\frac{h^4}{4!}f^{(4)}(x)-\frac{h^5}{5!}f^{(5)}(x)+\frac{h^6}{6!}f^{(6)}(\xi_2) \end{aligned}$$

Add the second to the first equation

$$f(x+h)+f(x-h) = 2f(x)+h^2f''(x)+\frac{h^4}{12}f^{(4)}(x)+\frac{h^6}{6!}\left[f^{(6)}(\xi_1)+f^{(6)}(\xi_2)\right]$$

Solve for $f''(x)$ we obtain

$$f''(x) = \frac{1}{h^2}\left[f(x+h)+f(x-h)\right]-\frac{2}{h^2}f(x)-\frac{h^2}{12}f^{(4)}(x)-\frac{1}{720}h^4\left[f^{(6)}(\xi_1)+f^{(6)}(\xi_2)\right] \quad (1A)$$

Now we do the same again, but by expanding around $x+2h$ and $x-2h$

$$\begin{aligned} f(x+2h) &= f(x)+2hf'(x)+\frac{(2h)^2}{2}f''(x)+\frac{(2h)^3}{3!}f^{(3)}(x)+\frac{(2h)^4}{4!}f^{(4)}(x)+\frac{(2h)^5}{5!}f^{(5)}(x)+\frac{(2h)^6}{6!}f^{(6)}(\xi_1) \\ f(x-2h) &= f(x)-2hf'(x)+\frac{(2h)^2}{2}f''(x)-\frac{(2h)^3}{3!}f^{(3)}(x)+\frac{(2h)^4}{4!}f^{(4)}(x)-\frac{(2h)^5}{5!}f^{(5)}(x)+\frac{(2h)^6}{6!}f^{(6)}(\xi_1) \end{aligned}$$

Add the second to the first equation

$$\begin{aligned} f(x+2h)+f(x-2h) &= 2f(x)+(2h)^2f''(x)+\frac{(2h)^4}{12}f^{(4)}(x)+\frac{(2h)^6}{6!}\left[f^{(6)}(\xi_1)+f^{(6)}(\xi_2)\right] \\ &= 2f(x)+4h^2f''(x)+\frac{4}{3}h^4f^{(4)}(x)+\frac{4}{45}h^6\left[f^{(6)}(\xi_1)+f^{(6)}(\xi_2)\right] \end{aligned}$$

Solve for $f''(x)$ we obtain

$$f''(x) = \frac{1}{4h^2}\left[f(x+2h) + f(x-2h)\right] - \frac{1}{2h^2}f(x) - \frac{1}{3}h^2 f^{(4)}(x) - \frac{1}{45}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] \tag{2A}$$

We want to eliminate $f^{(4)}(x)$ from the above. So we multiply eq(1A) by 4 and subtract eq(2) from the result. So equation (1A) becomes

$$\begin{aligned} 4f''(x) &= 4\left(\frac{1}{h^2}\left[f(x+h) + f(x-h)\right] - \frac{2}{h^2}f(x) - \frac{h^2}{12}f^{(4)}(x) - \frac{1}{720}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right]\right) \\ &= \frac{4}{h^2}\left[f(x+h) + f(x-h)\right] - \frac{8}{h^2}f(x) - \frac{1}{3}h^2 f^{(4)}(x) - \frac{1}{180}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] \end{aligned} \tag{3A}$$

Now subtract (2A) from (3A) we obtain

$$\begin{aligned} 3f''(x) = {} & \frac{4}{h^2}\left[f(x+h) + f(x-h)\right] - \frac{8}{h^2}f(x) - \frac{1}{3}h^2 f^{(4)}(x) - \frac{1}{180}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] - \\ & \left(\frac{1}{4h^2}\left[f(x+2h) + f(x-2h)\right] - \frac{1}{2h^2}f(x) - \frac{1}{3}h^2 f^{(4)}(x) - \frac{1}{45}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right]\right) \end{aligned}$$

Hence

$$\begin{aligned} 3f''(x) = {} & \frac{4}{h^2}\left[f(x+h) + f(x-h)\right] - \frac{8}{h^2}f(x) - \frac{1}{180}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] - \\ & \frac{1}{4h^2}\left[f(x+2h) + f(x-2h)\right] + \frac{1}{2h^2}f(x) + \frac{1}{45}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] \end{aligned}$$

$$\begin{aligned} f''(x) = {} & \frac{4}{3h^2}\left[f(x+h) + f(x-h)\right] - \frac{1}{12h^2}\left[f(x+2h) + f(x-2h)\right] - \frac{8}{3h^2}f(x) + \frac{1}{6h^2}f(x) - \\ & \frac{1}{3\times 180}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] + \frac{1}{3\times 45}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] \end{aligned}$$

$$\begin{aligned} &= \frac{1}{12h^2}\left(16\left[f(x+h) + f(x-h)\right] - \left[f(x+2h) + f(x-2h)\right] - 32f(x) + 2f(x)\right) + \frac{1}{180}h^4\left[f^{(6)}(\xi_1) + f^{(\right. \\ &= \frac{1}{12h^2}\left(16f(x+h) + 16f(x-h) - f(x+2h) - f(x-2h) - 30f(x)\right) + \frac{1}{180}h^4\left[f^{(6)}(\xi)\right] \end{aligned}$$

Rearrange terms to make it look as in the textbook

$$f''(x) = \frac{1}{12h^2}\left(-f\left(x+2h\right)+16f\left(x+h\right)-30f\left(x\right)+16f\left(x-h\right)-f\left(x-2h\right)\right)+\frac{1}{180}h^4\left[f^{(6)}\left(\xi\right)\right] \tag{4A}$$

Hence from equation $(4A)$ we see that the error is $O\left(h^4\right)$ as required to show.

Hence

$$\boxed{f''(x) \approx \frac{1}{12h^2}\left(-f\left(x+2h\right)+16f\left(x+h\right)-30f\left(x\right)+16f\left(x-h\right)-f\left(x-2h\right)\right)}$$

4.13.2 Section 7.1, Problem 9

problem: Show that in Richardson extrapolation, $D\left(2,2\right)=\frac{16}{15}\psi\left(\frac{h}{2}\right)-\frac{1}{15}\psi\left(h\right)$

Solution:

$$D\left(n,k\right)=\frac{4^k}{4^k-1}D\left(n,k-1\right)-\frac{1}{4^k-1}D\left(n-1,k-1\right) \tag{1}$$

Now, when $n=2,k=2$, we obtain from (1)

$$\begin{aligned} D\left(2,2\right) &= \frac{4^2}{4^2-1}D\left(2,1\right)-\frac{1}{4^2-1}D\left(1,1\right) \\ &= \frac{16}{15}D\left(2,1\right)-\frac{1}{15}D\left(1,1\right) \end{aligned}$$

But since $D\left(1,1\right)=\psi(h), D\left(2,1\right)=\psi\left(\frac{h}{2}\right)$

$$D\left(2,2\right)=\frac{16}{15}\psi\left(\frac{h}{2}\right)-\frac{1}{15}\psi(h)$$

4.13.3 Section 7.1, Problem 14

problem: Using Taylor series, derive the error term for the approximation

$$f'(x) \approx \frac{1}{2h}\left[-3f(x) + 4f(x+h) - f(x+2h)\right]$$

answer:

expand around $x+h$

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(\xi_1) \\ f'(x) &= \frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi_1) \qquad (1) \end{aligned}$$

Now expand around $x+2h$

$$\begin{aligned} f(x+2h) &= f(x) + 2hf'(x) + 2h^2f''(x) + \frac{8h^3}{6}f'''(\xi_2) \\ f'(x) &= \frac{1}{2h}f(x+2h) - \frac{1}{2h}f(x) - hf''(x) - \frac{4h^2}{6}f'''(\xi_2) \qquad (2) \end{aligned}$$

Multiply (2) by $-\frac{1}{2}$ and add result to (1) we obtain

$$\begin{aligned} -\frac{1}{2}f'(x) + f'(x) = {} & -\frac{1}{2}\left(\frac{1}{2h}f(x+2h) - \frac{1}{2h}f(x) - hf''(x) - \frac{4h^2}{6}f'''(\xi_2)\right) + \\ & \left(\frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi_1)\right) \end{aligned}$$

$$\begin{aligned} \frac{1}{2}f'(x) &= \frac{-1}{4h}f(x+2h) + \frac{1}{4h}f(x) + \frac{h}{2}f''(x) + \frac{2h^2}{6}f'''(\xi_2) + \frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi \\ f'(x) &= \frac{-1}{2h}f(x+2h) + \frac{1}{2h}f(x) + hf''(x) + \frac{4h^2}{6}f'''(\xi_2) + \frac{2}{h}f(x+h) - \frac{2}{h}f(x) - hf''(x) - \frac{2h^2}{6}f'''(\xi \\ &= \left[\frac{-1}{2h}f(x+2h) + \frac{1}{2h}f(x) + hf''(x) + \frac{2}{h}f(x+h) - \frac{2}{h}f(x) - hf''(x)\right] - \frac{2h^2}{6}f'''(\xi_1) + \frac{4h^2}{6}f''' \\ &= \frac{1}{2h}\left[-f(x+2h) + f(x) + 4f(x+h) - 4f(x)\right] - \frac{h^2}{3}f'''(\xi_1) + \frac{2h^2}{3}f'''(\xi_2) \\ &= \frac{1}{2h}\left[-f(x+2h) - 3f(x) + 4f(x+h)\right] - h^2\left(\frac{1}{3}f'''(\xi_1) + \frac{2}{3}f'''(\xi_2)\right) \end{aligned}$$

Which is the equation we are asked to show.

From the above we see that the error term is given by

$$h^2\left(\frac{1}{3}f'''\left(\xi_1\right)+\frac{2}{3}f'''\left(\xi_2\right)\right)$$

Hence the error is $O\left(h^2\right)$

4.13.4 Section 7.1, Problem 16

problem: Using Taylor series, derive the error term for the approximation

$$f''(x)\approx\frac{1}{h^2}\left[f\left(x\right)-2f\left(x+h\right)+f\left(x+2h\right)\right]$$

Answer: expand around $x+h$

$$\begin{aligned}f\left(x+h\right)&=f(x)+hf'(x)+\frac{h^2}{2}f''(x)+\frac{h^3}{6}f'''\left(\xi_1\right)\\ \frac{h^2}{2}f''(x)&=f\left(x+h\right)-f(x)-hf'(x)-\frac{h^3}{6}f'''\left(\xi_1\right)\\ f''(x)&=\frac{2}{h^2}f\left(x+h\right)-\frac{2}{h^2}f(x)-\frac{2}{h}f'(x)-\frac{h}{3}f'''\left(\xi_1\right)\qquad(1)\end{aligned}$$

Now expand around $x+2h$

$$\begin{aligned}f\left(x+2h\right)&=f(x)+2hf'(x)+2h^2f''(x)+\frac{8h^3}{6}f'''\left(\xi_2\right)\\ 2h^2f''(x)&=f\left(x+2h\right)-f\left(x\right)-2hf'(x)-\frac{8h^3}{6}f'''\left(\xi_2\right)\\ f''(x)&=\frac{1}{2h^2}f\left(x+2h\right)-\frac{1}{2h^2}f(x)-\frac{1}{h}f'(x)-\frac{2h}{3}f'''\left(\xi_2\right)\qquad(2)\end{aligned}$$

Multiply (2) by -2 and add result to (1) we obtain

$$
\begin{aligned}
-2f''(x)+f''(x) &= -2\left(\frac{1}{2h^2}f\left(x+2h\right)-\frac{1}{2h^2}f\left(x\right)-\frac{1}{h}f'(x)-\frac{2h}{3}f'''\left(\xi_2\right)\right)\\
&\quad+\frac{2}{h^2}f\left(x+h\right)-\frac{2}{h^2}f(x)-\frac{2}{h}f'(x)-\frac{h}{3}f'''\left(\xi_1\right)\\
-f''(x) &= -\frac{1}{h^2}f\left(x+2h\right)+\frac{1}{h^2}f(x)+\frac{2}{h}f'(x)+\frac{4h}{3}f'''\left(\xi_2\right)+\frac{2}{h^2}f\left(x+h\right)-\frac{2}{h^2}f(x)-\frac{2}{h}f'(x)-\frac{h}{3}f'''\left(\xi\right.\\
f''(x) &= \frac{1}{h^2}f\left(x+2h\right)-\frac{1}{h^2}f(x)-\frac{2}{h}f'(x)-\frac{4h}{3}f'''\left(\xi_2\right)-\frac{2}{h^2}f\left(x+h\right)+\frac{2}{h^2}f(x)+\frac{2}{h}f'(x)+\frac{h}{3}f'''\left(\xi_1\right.\\
&= \frac{1}{h^2}f\left(x+2h\right)-\frac{1}{h^2}f(x)-\frac{2}{h^2}f\left(x+h\right)+\frac{2}{h^2}f(x)+\frac{h}{3}f'''\left(\xi_1\right)-\frac{4h}{3}f'''\left(\xi_2\right)\\
&= \frac{1}{h^2}\left(f\left(x+2h\right)-f(x)-2f\left(x+h\right)+2f(x)\right)+h\left(\frac{1}{3}f'''\left(\xi_1\right)-\frac{4}{3}f'''\left(\xi_2\right)\right)\\
&= \frac{1}{h^2}\left(f\left(x+2h\right)+f(x)-2f\left(x+h\right)\right)+h\left(\frac{1}{3}f'''\left(\xi_1\right)-\frac{4}{3}f'''\left(\xi_2\right)\right)
\end{aligned}
$$

Hence

$$
f''(x)\approx\frac{1}{h^2}\left(f\left(x+2h\right)+f(x)-2f\left(x+h\right)\right)
$$

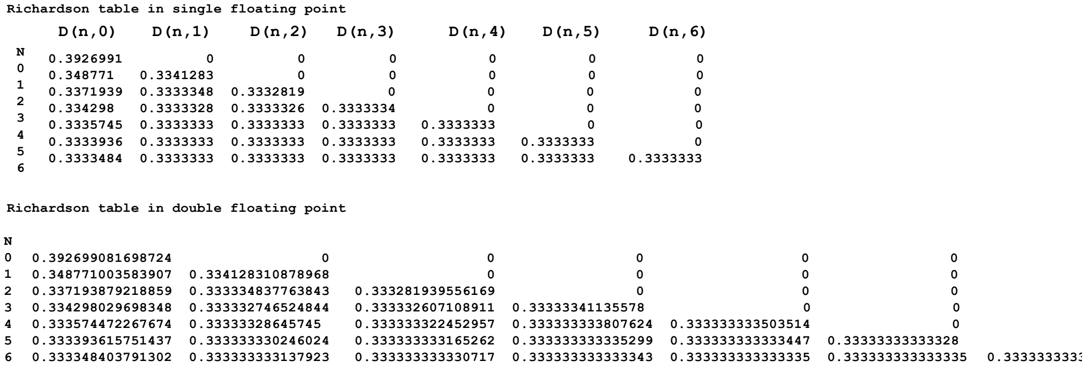
with the error term

$$
h\left(\frac{1}{3}f'''\left(\xi_1\right)-\frac{4}{3}f'''\left(\xi_2\right)\right)
$$

Hence the error is $O(h)$

4.13.5 Computer assignment 4/30/2007. Richardson Algorithm

This is the output



```
Richardson table in single floating point
      D(n,0)     D(n,1)     D(n,2)   D(n,3)       D(n,4)    D(n,5)      D(n,6)
 N
 0   0.3926991          0          0         0          0          0           0
 1   0.348771   0.3341283          0         0          0          0           0
 2   0.3371939  0.3333348  0.3332819         0          0          0           0
 3   0.334298   0.3333328  0.3333326  0.3333334         0          0           0
 4   0.3335745  0.3333333  0.3333333  0.3333333  0.3333333         0           0
 5   0.3333936  0.3333333  0.3333333  0.3333333  0.3333333  0.3333333          0
 6   0.3333484  0.3333333  0.3333333  0.3333333  0.3333333  0.3333333   0.3333333

Richardson table in double floating point

N
0  0.392699081698724                 0                 0                 0                 0                 0
1  0.348771003583907  0.334128310878968                 0                 0                 0                 0
2  0.337193879218859  0.333334837763843  0.333281939556169                 0                 0                 0
3  0.334298029698348  0.333332746524844  0.333332607108911  0.33333341135578                  0                 0
4  0.335574472267674  0.33333328645745   0.333333322452957  0.333333333807624  0.333333333503514                 0
5  0.333393615751437  0.333333330246024  0.333333333165262  0.333333333335299  0.333333333333447  0.33333333333328
6  0.333348403791302  0.333333333137923  0.333333333330717  0.333333333333343  0.333333333333335  0.333333333333335  0.333333333
```

Figure 4.7: Table output

This is the source code

```
%script to test nma_richardson
%Nasser Abbasi

h=1
x=sqrt(2);
f=@(x)atan(x);
M=6;

%first compute in single prcesion
 D=nma_richardson(h,x,f,M,0);
 format long g;
 fprintf('Richardson table in single floating point\n');
 D

 %Now do it in double prcesion
 D=nma_richardson(h,x,f,M,1);
 format long g;
 fprintf('Richardson table in double floating point\n');
 D
```

```
function D=nma_richardson(h,x,f,M,doubleFlag)
%function D=nma_richardson(h,x,f,M,doubleFlag)
%
%INPUT:
% h:  spacing for numerical differentiation
% x:  where to find diff
% f:  the function whos derivative we are finding
% M:  how big a richardson table to make
% doubleFlag: 0 to do it in single floating point
%             or 1 to do it in double floating


% MATH 501, CSUF, spring 2007
% computer assignment 4/30/2007
% Richardson extrapolation
%
%Nasser Abbasi, May 5,2007

if doubleFlag
    D=zeros(M+1,M+1);
else
    D=zeros(M+1,M+1,'single');
end

for n=1:M+1
    D(n,1)=phi(h/(2^(n-1)),x,f);
```

```
end

for k=2:M+1
    for n=k:M+1
        D(n,k)=D(n,k-1)+(D(n,k-1)-D(n-1,k-1))/(4^(k-1)-1);
    end
end
end


function r=phi(h,x,f)
r=1/(2*h)*(f(x+h)-f(x-h));
end
```

4.13.6 Computer assignment 5/2/2007. Midpoint,Trapezoid and Simpson

4.13.6.1 Conclusion

This table summarizes the results of the 3 methods

Method	RESULTS
Simpson	Error term $\frac{1}{180}(b-a)h^4\ \max\left\lvert f^{(4)}(\xi)\right\rvert$
	$I=\int_a^b f(x)dx \approx \frac{h}{3}\left(f(x_0)+2\sum_{i=2}^{N/2} f(x_{2i-2})+4\sum_{i=1}^{N/2} f(x_{2i-1})+f(x_N)\right)$
	Intervals needed: 900
	long format print of numerical integration: 90.379254649757272
Midpoint	Error term $\frac{1}{24}(b-a)h^2\max\left\lvert f^{(2)}(\xi)\right\rvert$
	$\int_a^b f(x)dx \approx h\sum_{i=1}^{N-1} f\left(\frac{x_{i+1}+x_i}{2}\right)$ note: N here is number of points
	Intervals needed: $174,285$
	long format print of numerical integration: 90.379254649446878
Trapezoid	Error term $\frac{1}{12}(b-a)h^2\max\left\lvert f^{(2)}(\xi)\right\rvert$
	$h\left(\frac{f(x_1)}{2}+\sum_{i=2}^{N-1} f(x_i)+\frac{f(x_N)}{2}\right)$ note: N here is number of points
	Intervals needed: $246,476$
	long format print of numerical integration: 90.379254649958952

4.13.6.2 Simpson

The error term in simpson is $\frac{1}{180}(b-a)h^4\ \max\left\lvert f^{(4)}(\xi)\right\rvert$ for some ξ between b,a. Since we want to limit the maximum error, we look to find where $f(\xi)$ is Max.

The function is $x\ln(x)$, hence $f^{(4)}(x)=\frac{2}{x^3}$ and this is maximum when x is smallest. Hence the maximum will occur at the lower end of the range, which is $x=1$ in this example.

Now we find the number of intervals N from solving $\frac{1}{180}(b-a)h^4\ \max\left\lvert f^{(4)}(\xi)\right\rvert < 10^{-9}$ where 10^{-9} is the error we are asked to limit our computation error to be below.

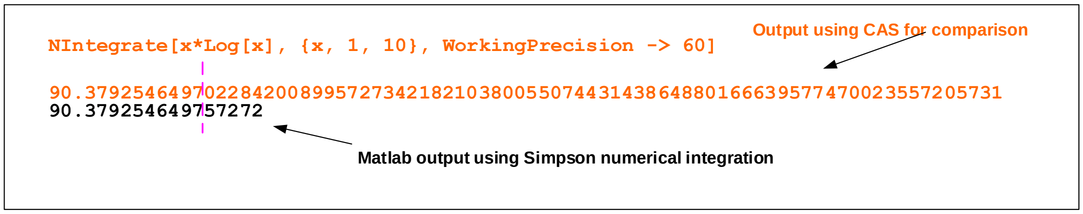
Next, we solve for h from the above. Knowing h, we find N which is the number of intervals. Next, we make sure N is even number by adjusting it if needed. We need to have even number of intervals Next we apply the simpson integration formula which is

$$I=\int_a^b f(x)dx \approx \frac{h}{3}\left(f(x_0)+2\sum_{i=2}^{N/2} f(x_{2i-2})+4\sum_{i=1}^{N/2} f(x_{2i-1})+f(x_N)\right)$$

In the above N is the number of intervals. Not to be confused with the following 2 algorithms below, where I used N to be number of points. For simpson, it was easier to stick with N being number of intervals.

The matlab implementation uses a vectorized version for speed.

To verify that the correct answer is obtain, it was compared with the output from a computer algebra system which uses an arbitrary large number of correct decimal points. The Matlab output was aligned against the CAS output and the digits verified to be correct to 9 decimal places are required.



```
NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]

90.3792546497022842008995727342182103800550744314386488016663957747002355720573 1
90.379254649757272
```

Output using CAS for comparison

Matlab output using Simpson numerical integration

Figure 4.8: Result

Source code:

```
function nma_simpson_math_501
%
%Math 501, CSUF, spring 2007
%Computer assignment 5/2/2007
%Nasser Abbasi

%For reference, this is exact answer for 60 decimal places
%NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]
%90.37925464970228420089957273421821038005507443143864880166639577470023557205731`60.

%

a = 1;
b = 10;
%maxError = 10^(-9);

%(2/x^3) is  d^4/dx^4 (x log(x))
%so max error will be when x is smallest, i.e. at a=1
I4      = abs(2/a^3);
errTerm = 1/180 * (b-a) * I4;
h       = maxError /errTerm;
h       = h^(1/4);
N       = ceil((b-a)/h);  % N is number of intervals

%N isnumber of intervals it needs to be EVEN number of intervals
if mod(N,2)==1
```

```
   N = N+1;
end

h = (b-a)/N;  %update h since we rounded up above.
fprintf('Simposon: Number of intervals needed is %d\n',N);

x = linspace(a,b,N+1);
f = @(x) x.*log(x);   %the function to integrate

%vectorized solution
I = f(x(1)) + 2*sum(f(x(3:2:end-2))) + 4*sum(f(x(2:2:end-1))) + f(x(end));
I = (h/3)*I;

fprintf('answer is'); format long;  I
```

4.13.6.3 Midpoint

The error term is $\frac{1}{24}(b-a)h^2\max\left|f^{(2)}(\xi)\right|$ for some ξ between b, a. Midterm is evaluated as follows

$$I=\int_a^b f(x)dx \approx h\sum_{i=1}^{N-1} f\left(\frac{x_{i+1}+x_i}{2}\right)$$

where N is the number of points. And I am using the Matlab convention for indexing, where the first point is x_1 and not x_0

We start by finding the number of intervals by solving for h from $\frac{1}{24}(b-a)h^2\max\left|f^{(2)}(\xi)\right| < 10^{-9}$ where 10^{-9} is the error we are asked to limit our computation error to be below.

The function is $x\ln(x)$, hence $f^{(2)}(x)=\frac{1}{x}$ which is maximum at $x=a$.

The matlab implementation is below with the output.

```
function nma_midpoint_math_501
%
%Math 501, CSUF, spring 2007
%Computer assignment 5/2/2007
%Nasser Abbasi

%For reference, this is exact answer for 60 decimal places
%NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]
%90.3792546497022842008995727342182103800550744314386488016663957747002355720573l`60.

%

a = 1;
b = 10;
```

```
maxError = 10^-9;

%d^2/dx^2 (x log(x)) is (1/x)
%so max error will be when x is smallest, i.e. at a=1
I2      = abs(1/a);
errTerm = 1/24 * (b-a) * I2;
h       = maxError /errTerm;
h       = sqrt(h);
N       = ceil((b-a)/h);
h       = (b-a)/N;  %update h since we rounded up above.
fprintf('Midpoint: Number of intervals needed is %d\n',N);

x     = linspace(a,b,N+1);
xbar  = (x(1:end-1)+x(2:end))/2;
f     = @(x) x.*log(x);   %the function to integrate

%vectorized solution
I = h*sum(f(xbar));

fprintf('answer is'); format long;  I
```

Output is

```
Midpoint: Number of intervals needed is 174285
answer is
I =

  90.379254649446878
```

4.13.6.4 Trapezoid numerical integration

The error term is $\frac{1}{12}(b-a)h^2 \max\left|f^{(2)}(\xi)\right|$ for some ξ between b, a. Trapezoid is evaluated as follows

$$h\left(\frac{f(x_1)}{2}+\sum_{i=2}^{n-1}f(x_i)+\frac{f(x_n)}{2}\right)$$

Where n is number of points, and I am using the Matlab indexing where x_1 is the first point, and not x_0, hence the last point is x_n

The following is the source code and the output

```
function nma_trap_math_501
%
%Math 501, CSUF, spring 2007
%Computer assignment 5/2/2007
```

```
%Nasser Abbasi

%For reference, this is exact answer for 60 decimal places
%NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]
%90.37925464970228420089957273421821038005507443143864880166639577470023557205731`60.

%

a = 1;
b = 10;
maxError = 10^-9;

%d^2/dx^2 (x log(x)) is (1/x)
%so max error will be when x is smallest, i.e. at a=1
I2      = abs(1/a);
errTerm = 1/12 * (b-a) * I2;
h       = maxError /errTerm;
h       = sqrt(h);
N       = ceil((b-a)/h);  % Number of intervals
h       = (b-a)/N;
fprintf('Trapezoid: Number of intervals needed is %d\n',N);

x     = linspace(a,b,N+1);
f     = @(x) x.*log(x);   %the function to integrate
fbar  = sum(f(x(2:end-1)));

%vectorized solution
I = h * ( f(x(1))/2 + fbar + f(x(end))/2 );

fprintf('answer is'); format long;  I
```

Output

```
Trapezoid: Number of intervals needed is 246476
answer is
I =

  90.379254649958952
```

4.13.7 source code

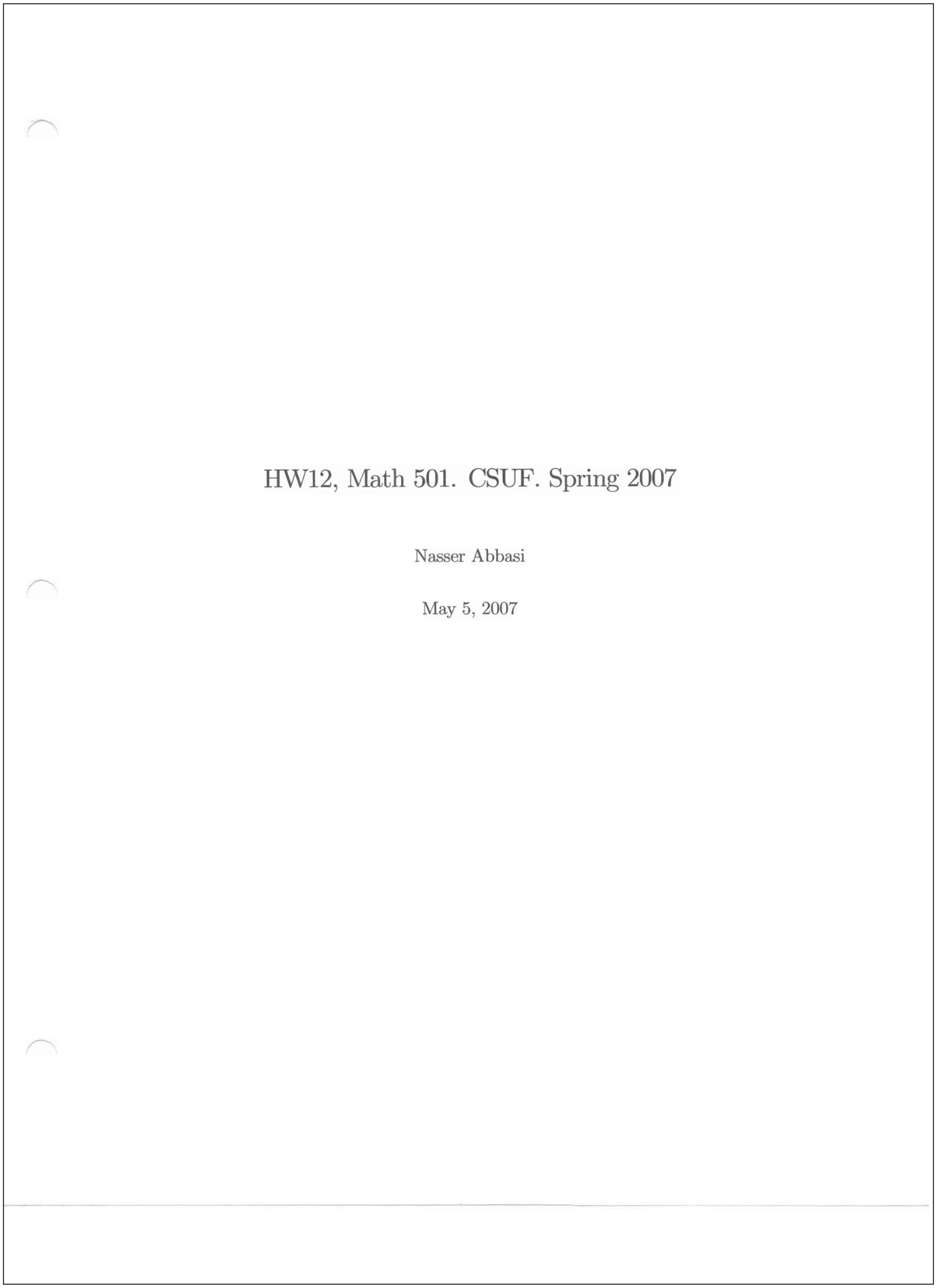
4.13.7.1 nma_compare.m

```
% Matlab code to illustrate the how the error changes in
% computing the derivative of arctan(x) at x=SQRT(2) as a function
% of changing h in Taylor approximation. Forcing Matlab to do the
% computation using 32 bits
% by Nasser Abbasi

h=single(1);
M=26;
X=single(sqrt(2));
f=@(x) single(atan(x));

F1=f(X);
S = zeros(26,6,'single');

for k=1:M
    F2=f(X+h);
    d=single(F2-F1);
    r=single(d/h);
    S(k,1)=k; S(k,2)=h; S(k,3)=F2; S(k,4)=F1; S(k,5)=d; S(k,6)=r;
    h=single(h/2);
end
format long g
S


% Matlab code to illustrate the how the error changes in
% computing the derivative of arctan(x) at x=SQRT(2) as a function
% of changing h in Taylor approximation. using Matlab default double
% floating point
% by Nasser Abbasi
clear all

h=1;
M=60;
X=sqrt(2);
f=@(x) atan(x);

F1=f(X);
S = zeros(26,6);

for k=1:M
    F2=f(X+h);
    d=F2-F1;
    r=d/h;
    S(k,1)=k; S(k,2)=h; S(k,3)=F2; S(k,4)=F1; S(k,5)=d; S(k,6)=r;
    h=h/2;
end
format long g
S
```

4.13.7.2 nma_trapezoidal.m

```
function I=nma_trapezoidal(func,from,to,nStrips)
%function r=nma_trapezoidal(f,from,to,nStrips)
%
% integrates a function using trapezoidal rule using
% specific number of strips.
%
% INPUT:
%   func : a string that repesents the function itself
%       for example 'x*sin(x)'. The independent variable
%       used in the string must be 'x' and no other letter.
%
%  from: lower limit
%  to  : upper limit
%  nStrips: number of strips to use
%
% OUTPUT
%   I : The integral.
%
% Author: Nasser Abbasi
% May 3, 2003

I=0;

if(nStrips<=0)
    error('number of strips must be > 0');
end

nPoints=nStrips+1;
X=linspace(from,to,nPoints);
h=X(2)-X(1);

for(i=1:length(X))
    x=X(i);
    f(i)=eval(func);
    if(i==1 | i==length(X) )
        I=I+f(i);
    else
        I=I+2*f(i);
    end
end

I=I/2;
I=I*h;
```

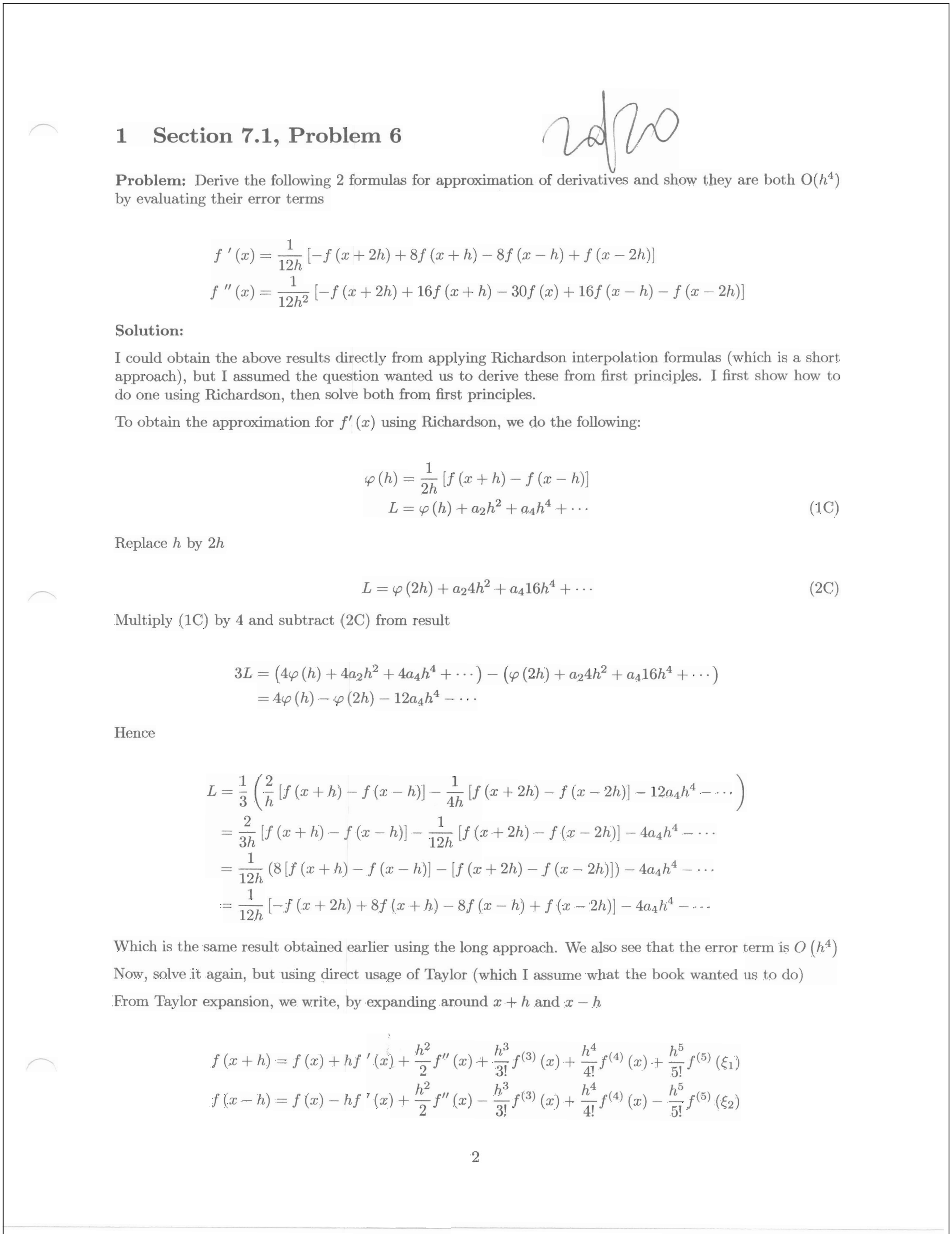
4.13.8 Graded



HW12, Math 501. CSUF. Spring 2007

Nasser Abbasi

May 5, 2007



1 Section 7.1, Problem 6

20/20

Problem: Derive the following 2 formulas for approximation of derivatives and show they are both O(h^4) by evaluating their error terms

$$f'(x) = \frac{1}{12h}\left[-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)\right]$$
$$f''(x) = \frac{1}{12h^2}\left[-f(x+2h) + 16f(x+h) - 30f(x) + 16f(x-h) - f(x-2h)\right]$$

Solution:

I could obtain the above results directly from applying Richardson interpolation formulas (which is a short approach), but I assumed the question wanted us to derive these from first principles. I first show how to do one using Richardson, then solve both from first principles.

To obtain the approximation for $f'(x)$ using Richardson, we do the following:

$$\begin{aligned} \varphi(h) &= \frac{1}{2h}\left[f(x+h) - f(x-h)\right] \\ L &= \varphi(h) + a_2h^2 + a_4h^4 + \cdots \end{aligned} \tag{1C}$$

Replace h by $2h$

$$L = \varphi(2h) + a_2 4h^2 + a_4 16h^4 + \cdots \tag{2C}$$

Multiply (1C) by 4 and subtract (2C) from result

$$\begin{aligned} 3L &= \left(4\varphi(h) + 4a_2h^2 + 4a_4h^4 + \cdots\right) - \left(\varphi(2h) + a_2 4h^2 + a_4 16h^4 + \cdots\right) \\ &= 4\varphi(h) - \varphi(2h) - 12a_4h^4 - \cdots \end{aligned}$$

Hence

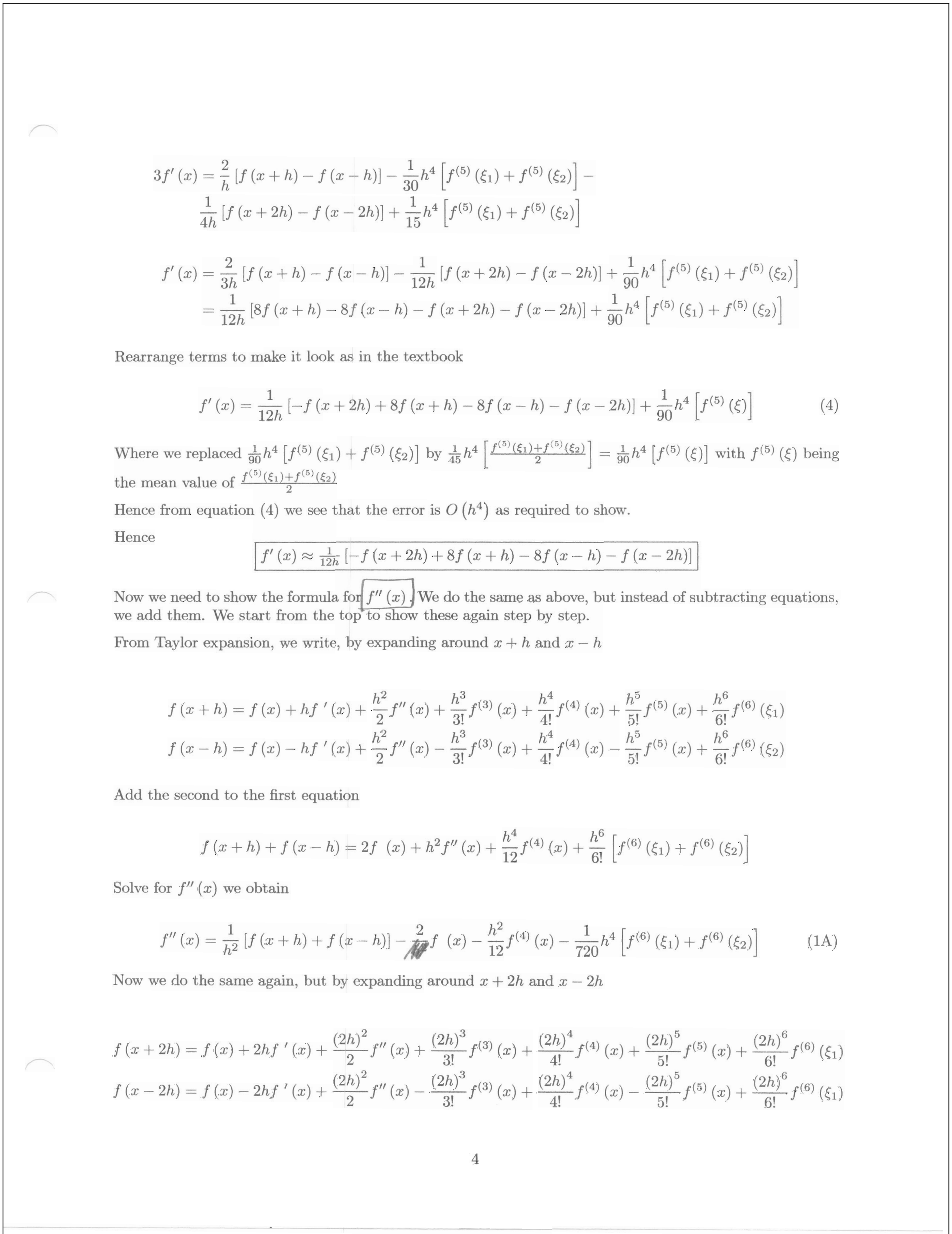
$$\begin{aligned} L &= \frac{1}{3}\left(\frac{2}{h}\left[f(x+h) - f(x-h)\right] - \frac{1}{4h}\left[f(x+2h) - f(x-2h)\right] - 12a_4h^4 - \cdots\right) \\ &= \frac{2}{3h}\left[f(x+h) - f(x-h)\right] - \frac{1}{12h}\left[f(x+2h) - f(x-2h)\right] - 4a_4h^4 - \cdots \\ &= \frac{1}{12h}\left(8\left[f(x+h) - f(x-h)\right] - \left[f(x+2h) - f(x-2h)\right]\right) - 4a_4h^4 - \cdots \\ &= \frac{1}{12h}\left[-f(x+2h) + 8f(x+h) - 8f(x-h) + f(x-2h)\right] - 4a_4h^4 - \cdots \end{aligned}$$

Which is the same result obtained earlier using the long approach. We also see that the error term is $O\left(h^4\right)$

Now, solve it again, but using direct usage of Taylor (which I assume what the book wanted us to do)

From Taylor expansion, we write, by expanding around $x+h$ and $x-h$

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) + \frac{h^5}{5!}f^{(5)}(\xi_1) \\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) - \frac{h^5}{5!}f^{(5)}(\xi_2) \end{aligned}$$



$$3f'(x) = \frac{2}{h}\left[f(x+h) - f(x-h)\right] - \frac{1}{30}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right] - \frac{1}{4h}\left[f(x+2h) - f(x-2h)\right] + \frac{1}{15}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right]$$

$$\begin{aligned} f'(x) &= \frac{2}{3h}\left[f(x+h) - f(x-h)\right] - \frac{1}{12h}\left[f(x+2h) - f(x-2h)\right] + \frac{1}{90}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right] \\ &= \frac{1}{12h}\left[8f(x+h) - 8f(x-h) - f(x+2h) - f(x-2h)\right] + \frac{1}{90}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right] \end{aligned}$$

Rearrange terms to make it look as in the textbook

$$f'(x) = \frac{1}{12h}\left[-f(x+2h) + 8f(x+h) - 8f(x-h) - f(x-2h)\right] + \frac{1}{90}h^4\left[f^{(5)}(\xi)\right] \tag{4}$$

Where we replaced $\frac{1}{90}h^4\left[f^{(5)}(\xi_1) + f^{(5)}(\xi_2)\right]$ by $\frac{1}{45}h^4\left[\frac{f^{(5)}(\xi_1)+f^{(5)}(\xi_2)}{2}\right] = \frac{1}{90}h^4\left[f^{(5)}(\xi)\right]$ with $f^{(5)}(\xi)$ being the mean value of $\frac{f^{(5)}(\xi_1)+f^{(5)}(\xi_2)}{2}$

Hence from equation (4) we see that the error is $O\left(h^4\right)$ as required to show.

Hence $\boxed{f'(x) \approx \frac{1}{12h}\left[-f(x+2h) + 8f(x+h) - 8f(x-h) - f(x-2h)\right]}$

Now we need to show the formula for $\boxed{f''(x)}$. We do the same as above, but instead of subtracting equations, we add them. We start from the top to show these again step by step.

From Taylor expansion, we write, by expanding around $x+h$ and $x-h$

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) + \frac{h^5}{5!}f^{(5)}(x) + \frac{h^6}{6!}f^{(6)}(\xi_1) \\ f(x-h) &= f(x) - hf'(x) + \frac{h^2}{2}f''(x) - \frac{h^3}{3!}f^{(3)}(x) + \frac{h^4}{4!}f^{(4)}(x) - \frac{h^5}{5!}f^{(5)}(x) + \frac{h^6}{6!}f^{(6)}(\xi_2) \end{aligned}$$

Add the second to the first equation

$$f(x+h) + f(x-h) = 2f(x) + h^2f''(x) + \frac{h^4}{12}f^{(4)}(x) + \frac{h^6}{6!}\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right]$$

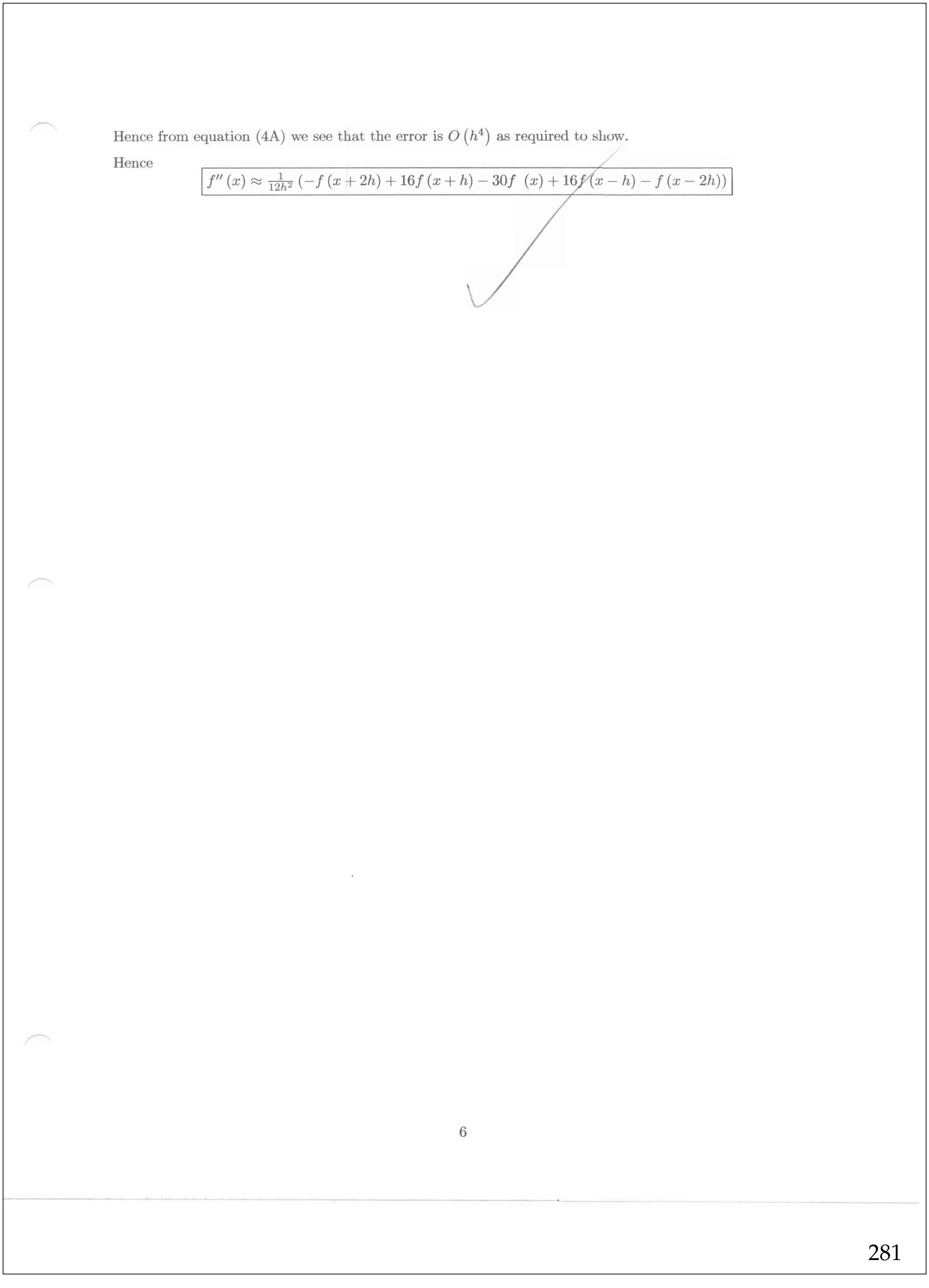
Solve for $f''(x)$ we obtain

$$f''(x) = \frac{1}{h^2}\left[f(x+h) + f(x-h)\right] - \frac{2}{h^2}f(x) - \frac{h^2}{12}f^{(4)}(x) - \frac{1}{720}h^4\left[f^{(6)}(\xi_1) + f^{(6)}(\xi_2)\right] \tag{1A}$$

Now we do the same again, but by expanding around $x+2h$ and $x-2h$

$$\begin{aligned} f(x+2h) &= f(x) + 2hf'(x) + \frac{(2h)^2}{2}f''(x) + \frac{(2h)^3}{3!}f^{(3)}(x) + \frac{(2h)^4}{4!}f^{(4)}(x) + \frac{(2h)^5}{5!}f^{(5)}(x) + \frac{(2h)^6}{6!}f^{(6)}(\xi_1) \\ f(x-2h) &= f(x) - 2hf'(x) + \frac{(2h)^2}{2}f''(x) - \frac{(2h)^3}{3!}f^{(3)}(x) + \frac{(2h)^4}{4!}f^{(4)}(x) - \frac{(2h)^5}{5!}f^{(5)}(x) + \frac{(2h)^6}{6!}f^{(6)}(\xi_1) \end{aligned}$$

4

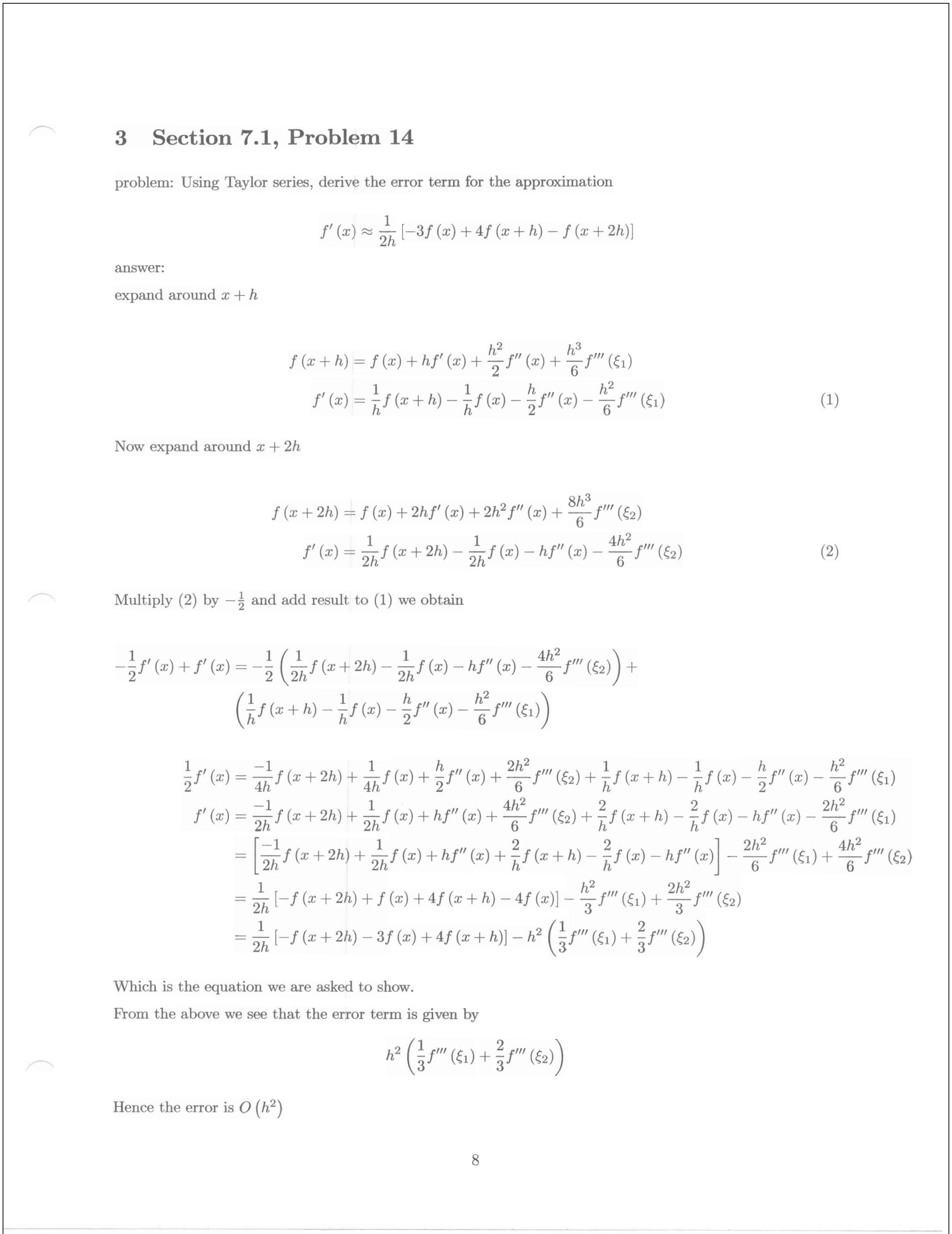


Hence from equation (4A) we see that the error is $O\left(h^4\right)$ as required to show.

Hence

$$\boxed{f''\left(x\right) \approx \frac{1}{12h^2}\left(-f\left(x+2h\right)+16f\left(x+h\right)-30f\ \left(x\right)+16f\left(x-h\right)-f\left(x-2h\right)\right)}$$

✓



3 Section 7.1, Problem 14

problem: Using Taylor series, derive the error term for the approximation

$$f'(x) \approx \frac{1}{2h}\left[-3f(x) + 4f(x+h) - f(x+2h)\right]$$

answer:

expand around $x+h$

$$\begin{aligned} f(x+h) &= f(x) + hf'(x) + \frac{h^2}{2}f''(x) + \frac{h^3}{6}f'''(\xi_1) \\ f'(x) &= \frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi_1) \qquad (1) \end{aligned}$$

Now expand around $x+2h$

$$\begin{aligned} f(x+2h) &= f(x) + 2hf'(x) + 2h^2f''(x) + \frac{8h^3}{6}f'''(\xi_2) \\ f'(x) &= \frac{1}{2h}f(x+2h) - \frac{1}{2h}f(x) - hf''(x) - \frac{4h^2}{6}f'''(\xi_2) \qquad (2) \end{aligned}$$

Multiply (2) by $-\frac{1}{2}$ and add result to (1) we obtain

$$\begin{aligned} -\frac{1}{2}f'(x) + f'(x) = -\frac{1}{2}&\left(\frac{1}{2h}f(x+2h) - \frac{1}{2h}f(x) - hf''(x) - \frac{4h^2}{6}f'''(\xi_2)\right) + \\ &\left(\frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi_1)\right) \end{aligned}$$

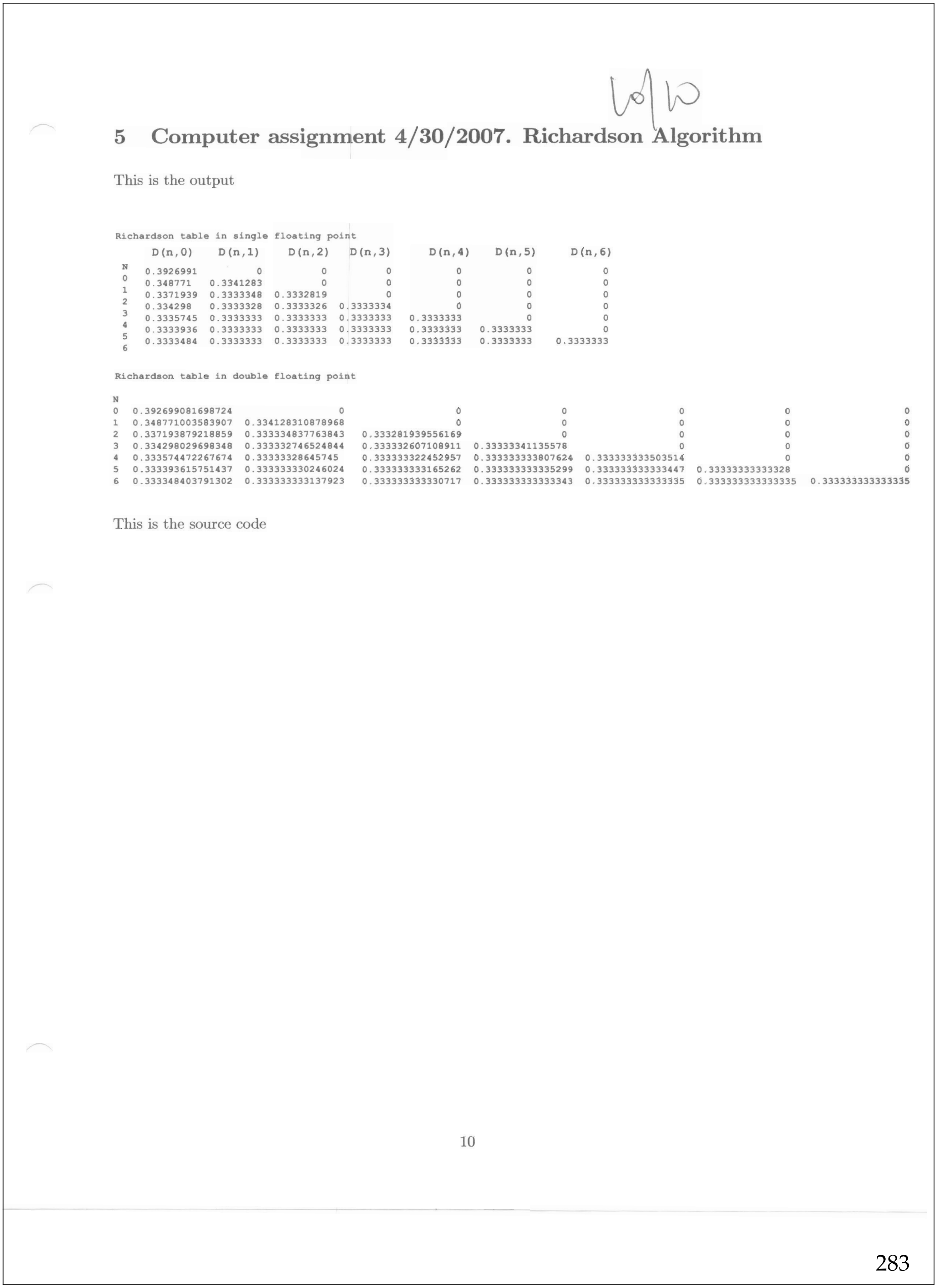
$$\begin{aligned} \frac{1}{2}f'(x) &= \frac{-1}{4h}f(x+2h) + \frac{1}{4h}f(x) + \frac{h}{2}f''(x) + \frac{2h^2}{6}f'''(\xi_2) + \frac{1}{h}f(x+h) - \frac{1}{h}f(x) - \frac{h}{2}f''(x) - \frac{h^2}{6}f'''(\xi_1) \\ f'(x) &= \frac{-1}{2h}f(x+2h) + \frac{1}{2h}f(x) + hf''(x) + \frac{4h^2}{6}f'''(\xi_2) + \frac{2}{h}f(x+h) - \frac{2}{h}f(x) - hf''(x) - \frac{2h^2}{6}f'''(\xi_1) \\ &= \left[\frac{-1}{2h}f(x+2h) + \frac{1}{2h}f(x) + hf''(x) + \frac{2}{h}f(x+h) - \frac{2}{h}f(x) - hf''(x)\right] - \frac{2h^2}{6}f'''(\xi_1) + \frac{4h^2}{6}f'''(\xi_2) \\ &= \frac{1}{2h}\left[-f(x+2h) + f(x) + 4f(x+h) - 4f(x)\right] - \frac{h^2}{3}f'''(\xi_1) + \frac{2h^2}{3}f'''(\xi_2) \\ &= \frac{1}{2h}\left[-f(x+2h) - 3f(x) + 4f(x+h)\right] - h^2\left(\frac{1}{3}f'''(\xi_1) + \frac{2}{3}f'''(\xi_2)\right) \end{aligned}$$

Which is the equation we are asked to show.

From the above we see that the error term is given by

$$h^2\left(\frac{1}{3}f'''(\xi_1) + \frac{2}{3}f'''(\xi_2)\right)$$

Hence the error is $O\left(h^2\right)$



10/10

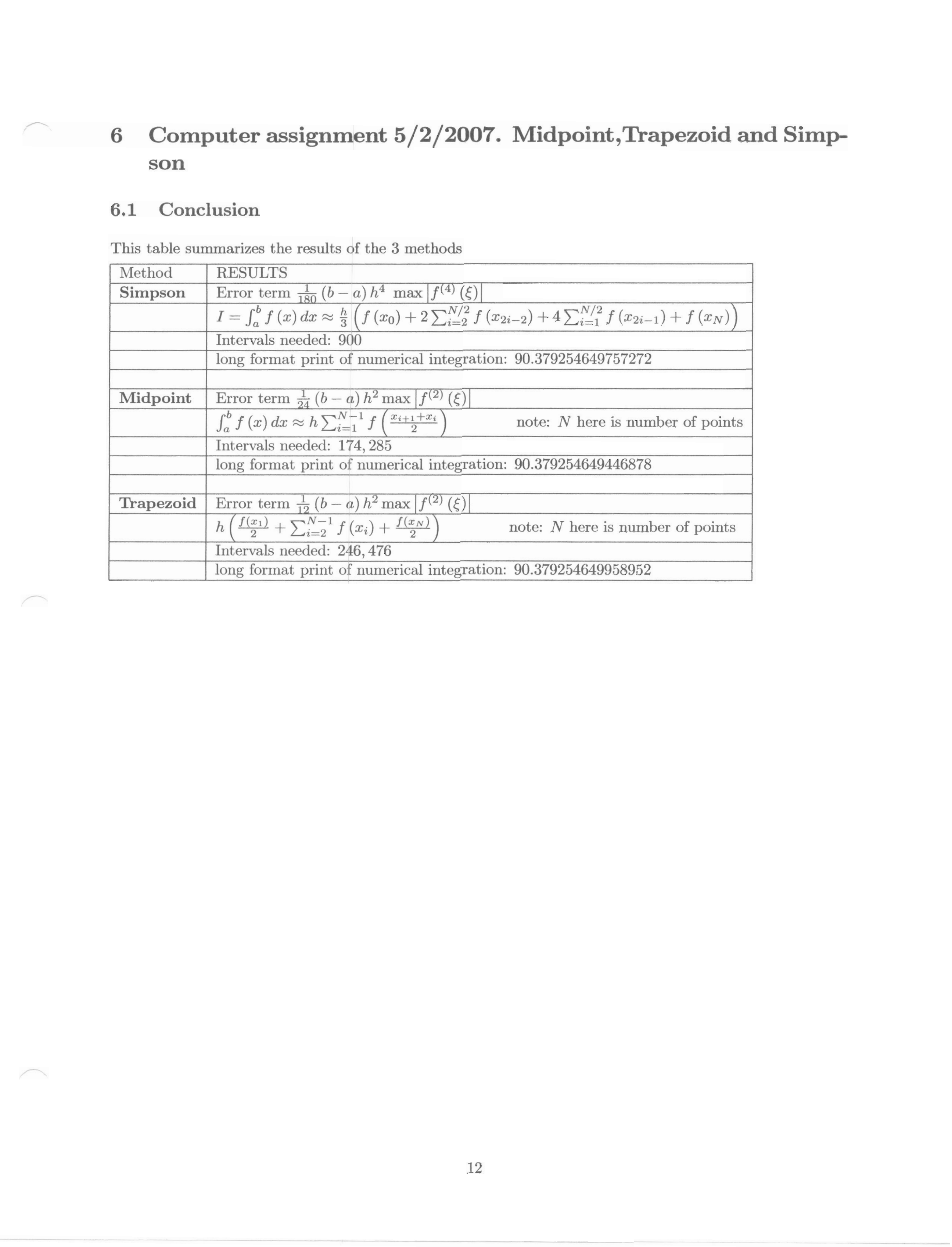
5 Computer assignment 4/30/2007. Richardson Algorithm

This is the output

```
Richardson table in single floating point
      D(n,0)     D(n,1)     D(n,2)    D(n,3)      D(n,4)     D(n,5)      D(n,6)
 N
 0   0.3926991          0          0          0          0          0          0
 1   0.348771   0.3341283          0          0          0          0          0
 2   0.3371939  0.3333348  0.3332819          0          0          0          0
 3   0.334298   0.3333328  0.3333326  0.3333334          0          0          0
 4   0.3335745  0.3333333  0.3333333  0.3333333  0.3333333          0          0
 5   0.3333936  0.3333333  0.3333333  0.3333333  0.3333333  0.3333333          0
 6   0.3333484  0.3333333  0.3333333  0.3333333  0.3333333  0.3333333  0.3333333
```

```
Richardson table in double floating point

N
0  0.392699081698724                   0                   0                  0                  0                  0                  0
1  0.348771003583907  0.334128310878968                    0                  0                  0                  0                  0
2  0.337193879218859  0.333334837763843  0.333281939556169                    0                  0                  0                  0
3  0.334298029698348  0.333332746524844  0.333332607108911  0.33333341135578                     0                  0                  0
4  0.333574472267674  0.33333328645745   0.333333322452957  0.333333333807624  0.333333333503514                    0                  0
5  0.333393615751437  0.333333330246024  0.333333333165262  0.333333333335299  0.333333333333447  0.33333333333328                     0
6  0.333348403791302  0.333333333137923  0.333333333330717  0.333333333333343  0.333333333333335  0.333333333333335  0.333333333333335
```

This is the source code

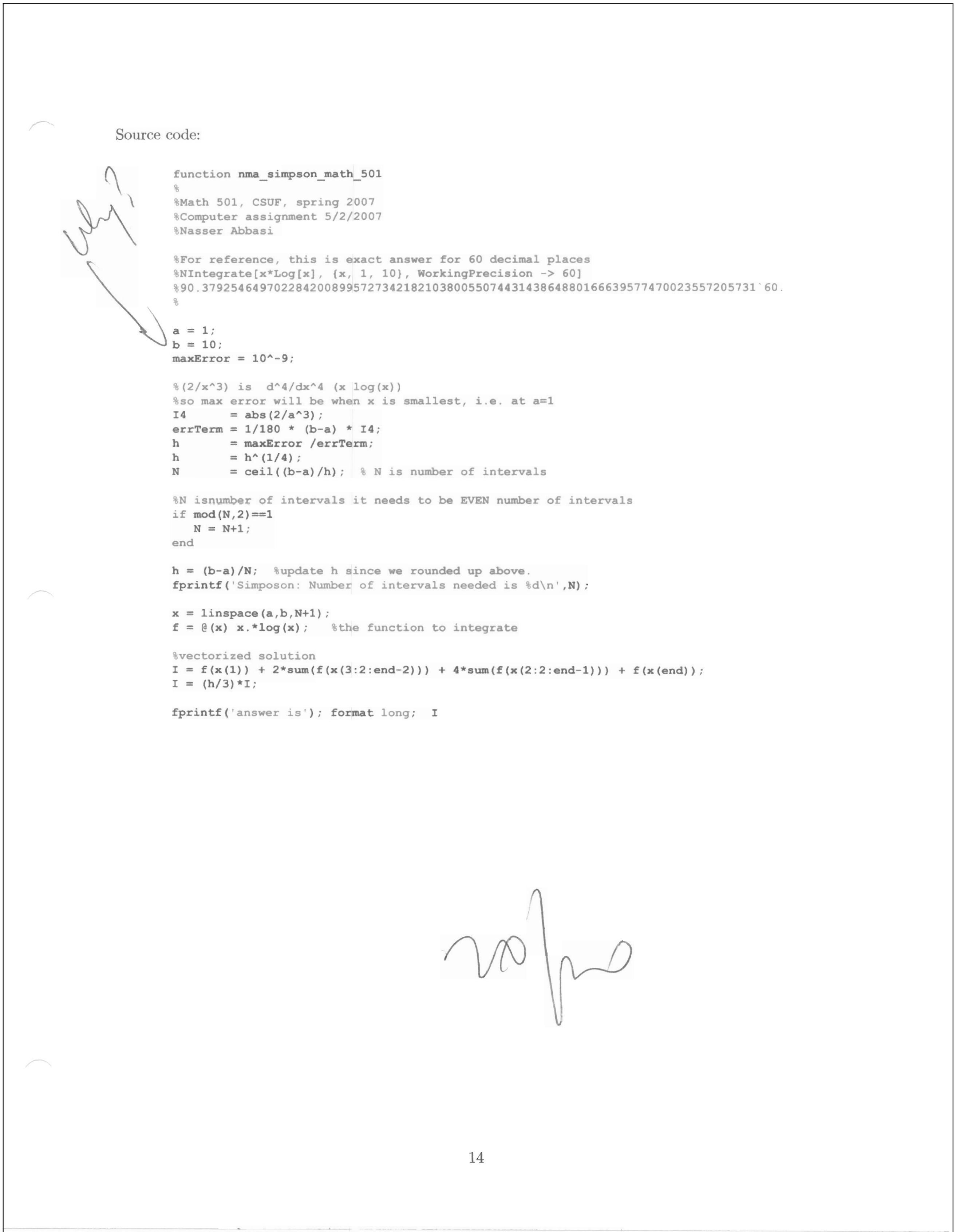


6 Computer assignment 5/2/2007. Midpoint,Trapezoid and Simpson

6.1 Conclusion

This table summarizes the results of the 3 methods

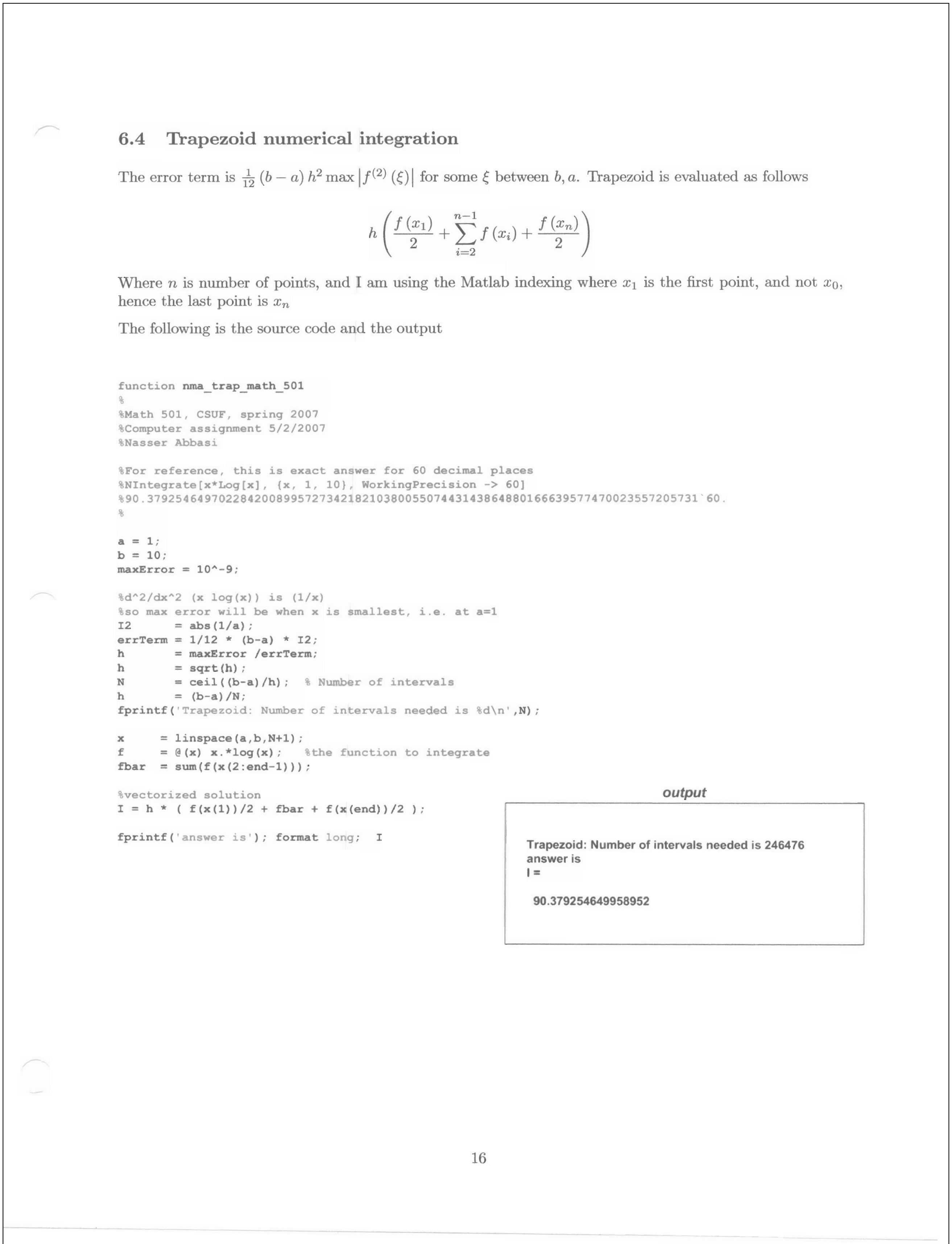
Method	RESULTS
Simpson	Error term $\frac{1}{180}\left(b-a\right)h^{4}\ \max\left\vert f^{\left(4\right)}\left(\xi\right)\right\vert$
	$I=\int_{a}^{b}f\left(x\right)dx\approx\frac{h}{3}\left(f\left(x_{0}\right)+2\sum_{i=2}^{N/2}f\left(x_{2i-2}\right)+4\sum_{i=1}^{N/2}f\left(x_{2i-1}\right)+f\left(x_{N}\right)\right)$
	Intervals needed: 900
	long format print of numerical integration: 90.379254649757272
Midpoint	Error term $\frac{1}{24}\left(b-a\right)h^{2}\max\left\vert f^{\left(2\right)}\left(\xi\right)\right\vert$
	$\int_{a}^{b}f\left(x\right)dx\approx h\sum_{i=1}^{N-1}f\left(\frac{x_{i+1}+x_{i}}{2}\right)$ note: N here is number of points
	Intervals needed: $174,285$
	long format print of numerical integration: 90.379254649446878
Trapezoid	Error term $\frac{1}{12}\left(b-a\right)h^{2}\max\left\vert f^{\left(2\right)}\left(\xi\right)\right\vert$
	$h\left(\frac{f\left(x_{1}\right)}{2}+\sum_{i=2}^{N-1}f\left(x_{i}\right)+\frac{f\left(x_{N}\right)}{2}\right)$ note: N here is number of points
	Intervals needed: $246,476$
	long format print of numerical integration: 90.379254649958952



Source code:

why?

```
function nma_simpson_math_501
%
%Math 501, CSUF, spring 2007
%Computer assignment 5/2/2007
%Nasser Abbasi

%For reference, this is exact answer for 60 decimal places
%NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]
%90.37925464970228420089957273421821038005507443143864880166639577470023557205731`60.
%

a = 1;
b = 10;
maxError = 10^-9;

%(2/x^3) is  d^4/dx^4 (x log(x))
%so max error will be when x is smallest, i.e. at a=1
I4      = abs(2/a^3);
errTerm = 1/180 * (b-a) * I4;
h       = maxError /errTerm;
h       = h^(1/4);
N       = ceil((b-a)/h);  % N is number of intervals

%N isnumber of intervals it needs to be EVEN number of intervals
if mod(N,2)==1
   N = N+1;
end

h = (b-a)/N;  %update h since we rounded up above.
fprintf('Simposon: Number of intervals needed is %d\n',N);

x = linspace(a,b,N+1);
f = @(x) x.*log(x);   %the function to integrate

%vectorized solution
I = f(x(1)) + 2*sum(f(x(3:2:end-2))) + 4*sum(f(x(2:2:end-1))) + f(x(end));
I = (h/3)*I;

fprintf('answer is'); format long;  I
```



6.4 Trapezoid numerical integration

The error term is $\frac{1}{12}(b-a)h^2 \max\left|f^{(2)}(\xi)\right|$ for some ξ between b, a. Trapezoid is evaluated as follows

$$h\left(\frac{f(x_1)}{2}+\sum_{i=2}^{n-1} f(x_i)+\frac{f(x_n)}{2}\right)$$

Where n is number of points, and I am using the Matlab indexing where x_1 is the first point, and not x_0, hence the last point is x_n

The following is the source code and the output

```
function nma_trap_math_501
%
%Math 501, CSUF, spring 2007
%Computer assignment 5/2/2007
%Nasser Abbasi

%For reference, this is exact answer for 60 decimal places
%NIntegrate[x*Log[x], {x, 1, 10}, WorkingPrecision -> 60]
%90.37925464970228420089957273421821038005507443143864880166639577470023557205731`60.
%

a = 1;
b = 10;
maxError = 10^-9;

%d^2/dx^2 (x log(x)) is (1/x)
%so max error will be when x is smallest, i.e. at a=1
I2      = abs(1/a);
errTerm = 1/12 * (b-a) * I2;
h       = maxError /errTerm;
h       = sqrt(h);
N       = ceil((b-a)/h);  % Number of intervals
h       = (b-a)/N;
fprintf('Trapezoid: Number of intervals needed is %d\n',N);

x     = linspace(a,b,N+1);
f     = @(x) x.*log(x);   %the function to integrate
fbar  = sum(f(x(2:end-1)));

%vectorized solution
I = h * ( f(x(1))/2 + fbar + f(x(end))/2 );

fprintf('answer is'); format long;  I
```

output

```
Trapezoid: Number of intervals needed is 246476
answer is
I =

 90.379254649958952
```

16