problem 1.

(a).

Let $\alpha = \alpha^2 - 2$, then we can find that there are two fixed points, one is -1 and another is 2.
Note: A fixed point α is stable if the absolute value of the derivative of the iteration function at α is strictly less than 1 and is unstable if the absolute value of the derivative of the iteration function at α is strictly greater than 1.

$$\phi'(x) = 2x$$
$$\implies |\phi'(-1)| = 2 > 1 \text{ and } |\phi'(2)| = 4 > 1.$$

Therefore, this iteration doesn't converge.

(b).

Let $\alpha = -\sqrt{\alpha + 2}$, then we can find that there are only one fixed point, which is 2.

$$\phi'(x) = -\frac{1}{2\sqrt{x+2}} \implies |\phi'(2)| = \frac{1}{2\sqrt{3}} < 1.$$

So we can say that 2 is a stable fixed point. Then we sketch the Cobweb diagram by Geogebra.

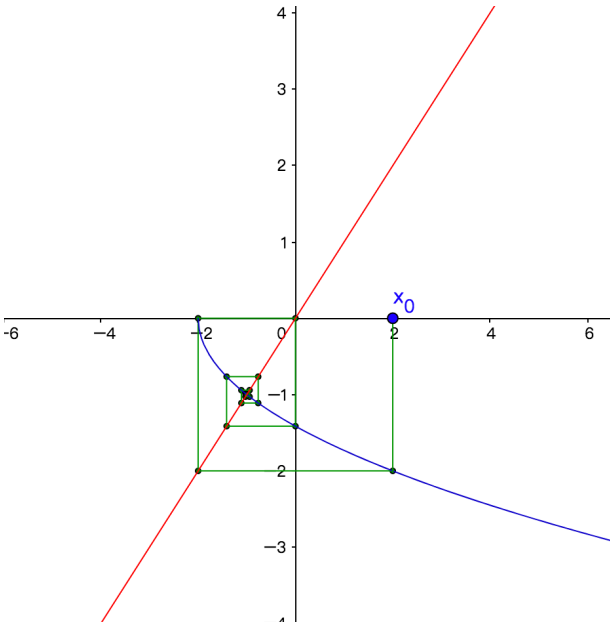

Figure 1: plot for $x_0 = 2$

By the Cobweb diagram, we can say when $x_0 \in (-2, 2)$ there exists $\lim x_n = -1$. Then we are going to find the order of convergence,

$$\lim_{n\to\infty} \frac{|-\sqrt{x_n+2}+1|}{|x_n+1|} \leq 2 < \infty$$

and

$$\lim_{n\to\infty} \frac{|-\sqrt{x_n+2}+1|}{|x_n+1|^\beta} = +\infty, \text{when } x_n = 1 \text{ and } \beta \geq 2,\ \beta \in \mathbb{N}.$$

Therefore, we can say the order of convergence is 1 and it's linearly convergence.

(c).

Let $\alpha = \alpha - 2 + \frac{\alpha}{\alpha - 1}$, the only fixed point is 2.

$$\phi'(x_n) = \frac{x^2 - 2x}{(x-1)^2}$$

$$\implies \phi'(2) = 0 < 1.$$

So we can say that 2 is a stable fixed point. Then we sketch the Cobweb diagram by Geogebra.

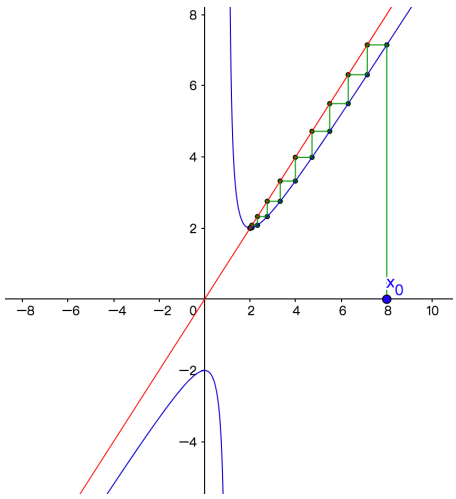

Figure 2: plot for $x_0 = 8$

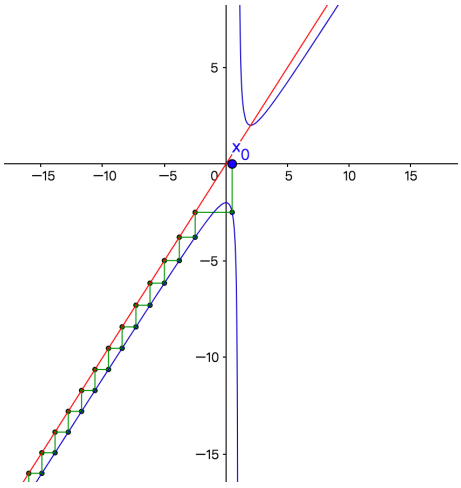

Figure 3: plot for $x_0 = 0.5$

By the Cobweb diagram, we can say when $x_0 \in (1, +\infty)$ there exists $\lim x_n = 2$. Then we are going to find the order of convergence,

$$\lim_{n\to\infty} \frac{|x_n - 4 + \frac{x_n}{x_n - 1}|}{|x_n - 2|} \leq \lambda < +\infty$$

and

$$\lim_{n\to\infty} \frac{|x_n - 4 + \frac{x_n}{x_n - 1}|}{|x_n - 2|^2} \leq \lambda < +\infty,$$

also we find that if $\beta \geq 2, \beta \in \mathbb{N}$,

$$\lim_{n\to\infty} \frac{|x_n - 4 + \frac{x_n}{x_n - 1}|}{|x_n - 2|^\beta} = +\infty.$$

Therefore, we can say the order of convergence is 2 and it's quadratically convergence.

problem 2.

(a).

We choose the Newton–Raphson method as our quadratically convergent method and the Relaxation method as the linearly one.

N-R Method: $x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$.
In this problem $\phi(x_n) = x_n - \frac{e^{-x} - \sin x}{(-e^{-x} - \cos x)}$, and the smallest positive root $\alpha \approx 0.5885$. We sketch the Cobweb diagram and find that if $x_0 \in (-\infty, \delta)$ such $\delta \in (1.7, 1.75)$, then we have

$$\lim_{n\to\infty} x_n = \alpha.$$

N-R method is quadratically convergent, to prove this we just need to provide that $f'(x) \neq 0$ near α,

$$f'(x) = -e^{-x} - \cos x \implies f'(\alpha) \neq 0,$$

Hence, N-R method is quadratically convergent holds here.

Relaxation Method: $x_{n+1} = \phi(x_n) = x_n - \omega f(x_n)$. $\omega \neq 0$ which is a constant.
$\phi'(x_n) = 1 - \omega f'(x_n)$, we want $|\phi'(x_n) < 1|$, and assume that $f'(x) = -e^{-x} - \cos x < 0$,

$$\implies \frac{2}{(-e^{-x} - \cos x)} < \omega < 0 \implies \omega \in (-1, 0).$$

By sketching the Cobweb diagram, we can see there exists a range such that we pick x_0 in this range then $\lim_{n\to\infty} x_n = \alpha$, α is the smallest positive root.
For example we pick $\omega = -0.5$,

$$\phi(x_n) = x_n + 0.5 f(x_n) = x_n + 0.5(e^{-x_n} - \sin x_n),$$

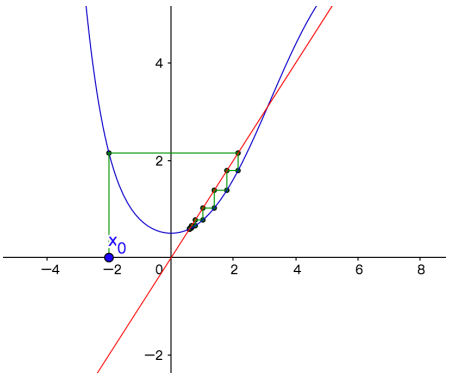

Figure 4: plot for $x_0 = -2$

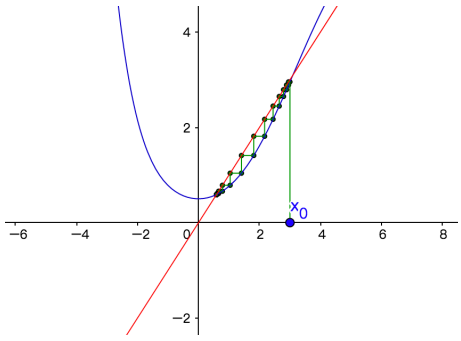

Figure 5: plot for $x_0 = 3$

Therefore, in this case $x_0 \in [-2, 3]$ approximately.
However, we need to show that the relaxation method is linearly convergent. Note that $\alpha \approx 0.5885, \omega \in (-1, 0)$.

$$\lim_{n\to+\infty} \frac{|x - \omega(e^{-x} - \sin x) - \alpha|}{|x - \alpha|} \leq \lambda < +\infty.$$

Therefore it's linearly convergent.

(b).

We choose the Newton–Raphson method as our quadratically convergent method and the Relaxation method as the linearly one.

N-R Method: $x_{n+1} = \phi(x_n) = x_n - \frac{f(x_n)}{f'(x_n)}$.
In this problem $\phi(x_n) = x_n - \frac{x - \cos x}{1 + \sin x}$, and the smallest positive root $\alpha \approx 0.739$. We sketch the Cobweb diagram and find that if $x_0 \in (-\infty, \delta)$ such $\delta \in [\approx -4.4, \approx 7.8]$, then we have

$$\lim_{n\to\infty} x_n = \alpha.$$

N-R method is quadratically convergent, to prove this we just need to provide that $f'(x) \neq 0$ near α,

$$f'(x) = 1 + \sin x \implies f'(\alpha) \neq 0,$$

Hence, N-R method is quadratically convergent holds here.

Relaxation Method: $x_{n+1} = \phi(x_n) = x_n - \omega f(x_n)$. $\omega \neq 0$ which is a constant.
$\phi'(x_n) = 1 - \omega f'(x_n)$, we want $|\phi'(x_n) < 1|$, and assume that $f'(x) = 1 + \sin x > 0$,

$$\implies \frac{2}{1 + \sin x} > \omega > 0 \implies \omega \in (0, 1).$$

By sketching the Cobweb diagram, we can see there exists a range such that we pick x_0 in this range then $\lim_{n\to\infty} x_n = \alpha$, α is the smallest positive root.
For example we pick $\omega = 0.5$,

$$\phi(x_n) = x_n - 0.5 f(x_n) = x_n + 0.5(x - \cos x),$$

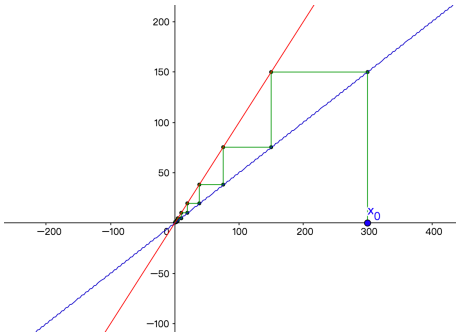

Figure 6: plot for $x_0 = 300$

Therefore, in this case $x_0 \in \mathbb{R}$ approximately.
However, we need to show that the relaxation method is linearly convergent. Note that $\alpha \approx 0.739, \omega \in (0, 1)$.

$$\lim_{n \to +\infty} \frac{|x - \omega(x - \cos x) - \alpha|}{|x - \alpha|} \leq \lambda < +\infty.$$

Therefore it's linearly convergent.

problem 3.

(a).

We do elimination by columns from left to right, rather than by rows from top to bottom. This is equal to post multiplication by an upper triangular matrix. We need to do $n - 1$ such operation to make A is lower triangular. We have the operations s.t.

$$\text{“}AU_1U_2 \cdots U_{n-2}U_{n-1} = L \implies A = LU_{n-1}U_{n-2} \cdots U_2U_1\text{”}.$$

Note that the inverse of an upper-triangular matrix is still upper triangular and multiplication of two upper-triangular matrix is also an upper-triangular matrix. Therefore, we have $A = LU$.

(b).

Gaussian Elimination is equal to pre-multiplication with $n - 1$ lower-triangular matrices. After the rescaling by a diagonal matrix D, the elimination can be written as

$$\text{“}L_{n-1}L_{n-2} \cdots L_2L_1AD = U \implies A = LUD^{-1}\text{”},$$

thus the unknown is rescaled by D^{-1}.

(c).

First, we know that $A = LU$, then we do additional operations, we use (a) then we have $U = D\tilde{U}$. Hence,

$$A = LD\tilde{U}.$$

Note that $\tilde{U}$ incorporate the additional column operations.

problem 4.

(a).

We separate our algorithm into two steps.

1. The first step is "inverse from LU factorization".

$$A^{-1} = (PLU)^{-1} = U^{-1}L^{-1}P^T,$$

this step costs $\frac{2}{3}n^3$ floating point multiplication/division operations.

2. The second step is solving "$AX = I$" by n equation s.t.

$$Ax_1 = e_1, \cdots, Ax_n = e_n,$$

x_i here is each column of X and e_i is the ith unit vector of size n, this step costs $2n^3$ floating point multiplication/division operations.
Therefore the sum of operations is $\frac{8}{3}n^3$, if we say that the algorithm is bounded by $Cn^3 + O(n^2)$ as $n \to +\infty$, the best value of C is $\frac{8}{3}$.

(b).

Note: The second step of our algorithm contains two parts, forward substitution and back substitution, each part have n^3 floating point multiplication/division operations.
We do a variant of our algorithm, taking the advantage of sparsity, and reduce the operations. When doing the both forward and back substitution for e_i, the first $i-1$ elements will be zero. The number of flops can then be reduced to

$$\sum_{i=1}^{n}\sum_{k=i}^{n} 2(n-k) \approx \frac{1}{3}n^3,$$

so the sum of the operations is

$$2 * \frac{1}{3}n^3 + \frac{2}{3}n^3 = \frac{4}{3}n^3.$$

If we say that the algorithm is bounded by $cn^3 + O(n^2)$ as $n \to +\infty$, with $c = \frac{4}{3}$ s.t. $c \sim \frac{C}{2}$.

(c).

(i). Counting the operations directly from the LU factorization. First, the LU factorization costs $\frac{2}{3}n^3$ floating point multiplication/division operations. For solving each system, it takes $2n^2$ floating point multiplication/division operations. There are m systems of equation, so the sum of operations is $\frac{2}{3}n^3 + 2mn^2$.
(ii). Counting the operations with a preliminary computation of A^{-1}. Producing the inverse requires $\frac{4}{3}n^3$, so the sum of operations is $\frac{4}{3}n^3 + 2mn^2$.

problem 5.

First we need to prove that $|\tilde{A} - A|$ is bounded in absolute value by an expression depending only on ϵ, n, m.

Proof. We pick a_i which is the i-th row of the matrix A, so

$$a_i x = y_i, i \in [1, m].$$

We can also find that

$$\widetilde{a_i} \otimes x = \tilde{y_i}, i \in [1, m]$$

using floating point arithmetic. Here we put all the errors in a_i s.t.

$$\begin{aligned}
&\widetilde{a_i} = (1+\delta)(1+\eta)a_i, |\delta|, |\eta| \leq \epsilon \\
&\implies \widetilde{a_{ij}} = (1+\delta)(1+\eta)a_{ij} \leq (1+\epsilon)^2 a_{ij}, |\delta|, |\eta| \leq \epsilon, i \in [1,m], j \in [1,n] \\
&\implies |\widetilde{a_{ij}} - a_{ij}| \leq |(1+\epsilon)^2 - 1||a_{ij}|, i \in [1,m], j \in [1,n]m.
\end{aligned}$$

Note that a_{ij} and $\widetilde{a_{ij}}$ are each element of the matrix A and $\widetilde{A}$ respectively. Hence we can bound $|\tilde{A} - A|$ in given condition. □

Then we need to show that matrix-multiplication is backward stable.

Proof. We show this statement by relative error measure.

$$\max_{ij} \frac{|\widetilde{a_{ij}} - a_{ij}|}{|a_{ij}|} \leq \max_{ij} \frac{|(1+\epsilon)^2 - 1||a_{ij}|}{|a_{ij}|} \leq (1+\epsilon)^2 - 1 \in O(\epsilon).$$

Therefore, matrix-multiplication is backward stable. □

Then we estimate the error $||\tilde{y} - y||$,

$$\tilde{y_i} - y_i = \tilde{a_i}x - a_i x = ((1+\delta)(1+\eta) - 1)y_i,$$

$$||\tilde{y_i} - y_i|| = \sqrt{\sum_{i=1}^{m} (\tilde{y_i} - y_i)^2} \leq \sqrt{\sum_{i=1}^{m} ((1+\epsilon)^2 - 1)y_i^2} \leq \sqrt{((1+\epsilon)^2 - 1)}||y|| = O(\epsilon)||y||.$$

Remark:

I have worked this assignment with David Knapik, Luke Steverango, Ralph Sarkis and Carl Perreault-Lafleur.