Consider backward & trapezoidal difference schemes for:

$$\frac{dy}{dt} = F(y,t) = \gamma y = (\lambda + i\omega) y \tag{1}$$

1. Estimate the order of accuracy for these two schemes

a. Backward:

$$\frac{dy}{dt} \simeq \frac{y(t_n) - y(t_n - \Delta t)}{\Delta t} = \frac{y(t_n) - y(t_n + (-\Delta t))}{\Delta t} \tag{2}$$

Taylor Expand:

$$y(t_n + (-\Delta t)) = y(t_n) + \left.\frac{dy}{dt}\right|_{t_n} (-\Delta t) + \frac{(-\Delta t)^2}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \text{H.O.T.} \tag{3}$$

move $y(t_n)$ over

$$y(t_n + (-\Delta t)) - y(t_n) = -\Delta t \left.\frac{dy}{dt}\right|_{t_n} + \frac{(\Delta t)^2}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \text{H.O.T.} \tag{4}$$

Divide by $-\Delta t$

$$\frac{y(t_n) - y(t_n - \Delta t)}{\Delta t} = \left.\frac{dy}{dt}\right|_{t_n} - \frac{\Delta t}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \text{H.O.T.} \tag{5}$$

$\therefore$ Backward-Euler has <u>First-order</u> accuracy: $O(\Delta t)$

b. Trapezoidal:

$$\frac{dy}{dt} \simeq \frac{y(t_{n+1}) - y(t_n)}{\Delta t} = \frac{1}{2}\left(\left.\frac{dy}{dt}\right|_{t_n} + \left.\frac{dy}{dt}\right|_{t_{n+1}} \right) \tag{6}$$

Taylor expansion of $\left.\frac{dy}{dt}\right|_{t_{n+1}}$:

$$\left.\frac{dy}{dt}\right|_{t_{n+1}} \simeq \left.\frac{dy}{dt}\right|_{t_n} + \Delta t \left.\frac{d^2y}{dt^2}\right|_{t_n} + \frac{(\Delta t)^2}{2} \frac{d^3y}{dt^3} + \text{H.O.T.} \tag{7}$$

From Forward Euler we know:

$$\frac{y(t_{n+1}) - y(t_n)}{\Delta t} = \left.\frac{dy}{dt}\right|_{t_n} + \frac{\Delta t}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \frac{(\Delta t)^2}{6} \left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T.} \tag{8}$$

Setting this equal to the RHS of eq. (6):

$$\frac{1}{2}\left(\left.\frac{dy}{dt}\right|_{t_n} + \left.\frac{dy}{dt}\right|_{t_{n+1}} \right) = \left.\frac{dy}{dt}\right|_{t_n} + \frac{\Delta t}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \frac{(\Delta t)^2}{6} \left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T.} \tag{9}$$

Plugging in expression from eq. (7) to the LHS of eq. (9):

$$\frac{1}{2}\left(\left.\frac{dy}{dt}\right|_{t_n} + \left.\frac{dy}{dt}\right|_{t_n} + \Delta t \left.\frac{d^2y}{dt^2}\right|_{t_n} + \frac{(\Delta t)^2}{2} \left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T} \right) = \left.\frac{dy}{dt}\right|_{t_n} + \frac{\Delta t}{2} \left.\frac{d^2y}{dt^2}\right|_{t_n} + \frac{(\Delta t)^2}{6} \left.\frac{d^3y}{dt^3}\right|_{t_n} \tag{10}$$

Now cancelling out like terms:

$$\frac{(\Delta t)^2}{4}\left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T.} = \frac{(\Delta t)^2}{6}\left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T.} \tag{11}$$

$$\therefore \boxed{\frac{y(t_{n+1}) - y(t_n)}{\Delta t} \approx \frac{(\Delta t)^2}{12}\left.\frac{d^3y}{dt^3}\right|_{t_n} + \text{H.O.T.}} \tag{12}$$

So trapezoidal approximation results in second-order accuracy! $\mathcal{O}((\Delta t)^2)$

7. Derive A-stability criteria for these two difference schemes & compare results with the figure in the file:

From eq. (2.21) in the book we know:

$$\frac{y_{n+1} - y_n}{\Delta t} = (1-\alpha) F(y_n, t_n) + \alpha F(y_{n+1}, t_{n+1}) \quad ; \quad \alpha = \begin{cases} 0 & \text{Forward Euler} \\ 1 & \text{Backward Euler} \\ \frac{1}{2} & \text{Trapezoidal} \end{cases} \tag{13}$$

Also, $$F(y_n) = \gamma y_n \tag{14}$$

Combining eqs. (13) & (14):

$$\frac{y_{n+1} - y_n}{\Delta t} = (1-\alpha)\gamma y_n + \alpha\gamma y_{n+1} \tag{15}$$

Now rearranging to get $\left|\frac{y_{n+1}}{y_n}\right| = |A|$:

$$y_{n+1} - y_n = (1-\alpha)\gamma y_n \Delta t + \alpha\gamma y_{n+1}\Delta t \tag{16}$$

$$y_{n+1} - \alpha\gamma y_{n+1}\Delta t = y_n + (1-\alpha)\gamma y_n \Delta t \tag{17}$$

$$y_{n+1}(1 - \alpha\gamma\Delta t) = y_n(1 + (1-\alpha)\gamma\Delta t) \tag{18}$$

$$\frac{y_{n+1}}{y_n} = \frac{1 + (1-\alpha)\gamma\Delta t}{1 - \alpha\gamma\Delta t} \tag{19}$$

which is eq. (2.22) from the book. Now A-stability requires $|A| \leq 1$, so first let's examine backward Euler with $\alpha = 1$:

$$\frac{y_{n+1}}{y_n} = \frac{1}{1 - \gamma\Delta t} \tag{20}$$

$$\frac{y_{n+1}}{y_n} = \frac{1}{1 - (\lambda + i\omega)\Delta t} \tag{21}$$

Rewriting the denominator as a singular complex number gives

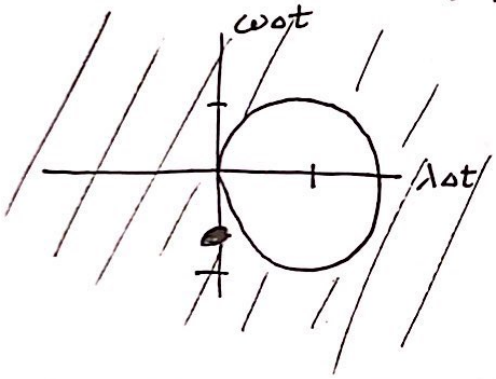
$$\frac{y_{n+1}}{y_n} = \frac{1}{(1+\lambda\Delta t) + i\omega\Delta t} \tag{22}$$

Since $|A| \leq 1$, and $|A| = \left|\frac{y_{n+1}}{y_n}\right|$, it can also be stated that $|A|^2 \leq 1$. Writing (22) in this form eliminates the imaginary component of the denominator:

$$|A|^2 = \left|\frac{y_{n+1}}{y_n}\right|^2 = \frac{1}{(1+\lambda\Delta t)^2 + (\omega\Delta t)^2} \leq 1 \tag{23}$$

$$\therefore \boxed{(1+\lambda\Delta t)^2 + (\omega\Delta t)^2 \geq 1} \tag{24}$$

The result gives $|A| \leq 1$ $\forall (\lambda\Delta t, \omega\Delta t)$ that lie outside of a circle of radius 1 centered at (1, 0)



Now examining trapezoidal in which $\alpha = \frac{1}{2}$ in eq. (19):

$$\frac{y_{n+1}}{y_n} = \frac{1 + \frac{1}{2}\gamma\Delta t}{1 - \frac{1}{2}\gamma\Delta t} \tag{25}$$

$$\frac{y_{n+1}}{y_n} = \frac{1 + \frac{1}{2}(\lambda + i\omega)\Delta t}{1 - \frac{1}{2}(\lambda + i\omega)\Delta t} = \frac{(1 + \frac{1}{2}\lambda\Delta t) + i\frac{1}{2}\omega\Delta t}{(1 - \frac{1}{2}\lambda\Delta t) - i\frac{1}{2}\omega\Delta t} \tag{26}$$

Again, evaluating $|A|^2$ to be $|A|^2 \leq 1$

$$\left|\frac{y_{n+1}}{y_n}\right|^2 = \frac{(1 + \frac{\lambda}{2}\Delta t)^2 + (\frac{\omega}{2}\Delta t)^2}{(1 - \frac{\lambda}{2}\Delta t)^2 + (\frac{\omega}{2}\Delta t)^2} \leq 1 \tag{27}$$

Since all terms in the denominator are positive, we can multiply the denominator to both sides to end up with:

$$(1 + \frac{\lambda}{2}\Delta t)^2 + (\frac{\omega}{2}\Delta t)^2 \leq (1 - \frac{\lambda}{2}\Delta t)^2 + (\frac{\omega}{2}\Delta t)^2 \tag{28}$$

Canceling the $(\frac{\omega}{2}\Delta t)^2$ terms,

$$\left(1+\frac{\lambda}{2}\Delta t\right)^2 \leq \left(1-\frac{\lambda}{2}\Delta t\right)^2 \tag{29}$$

$$\left(1+\frac{\lambda}{2}\Delta t\right)^2-\left(1-\frac{\lambda}{2}\Delta t\right)^2 \leq 0 \tag{30}$$

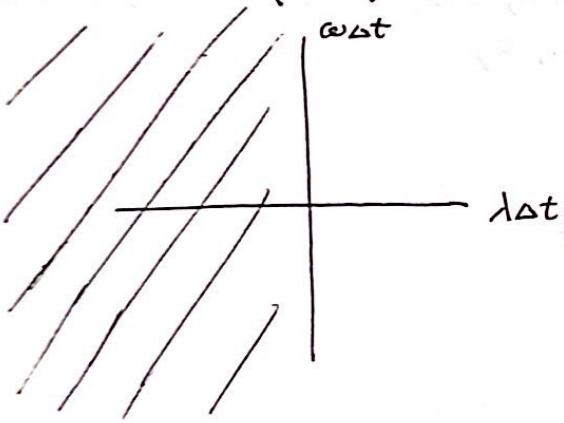
$$1+\lambda\Delta t+\frac{\lambda^2}{4}(\Delta t)^2-1+\lambda\Delta t-\frac{\lambda^2}{4}(\Delta t)^2 \leq 0 \tag{31}$$

Cancelling terms that are opposites:

$$2\lambda\Delta t \leq 0 \tag{32}$$

$$\therefore \boxed{\lambda\Delta t \leq 0} \tag{33}$$

This gives $|A| \leq 1 \ \forall \ \lambda\Delta t \leq 0$, which equates to the entire II + III quadrants of a graph of $\omega\Delta t(\lambda\Delta t)$:



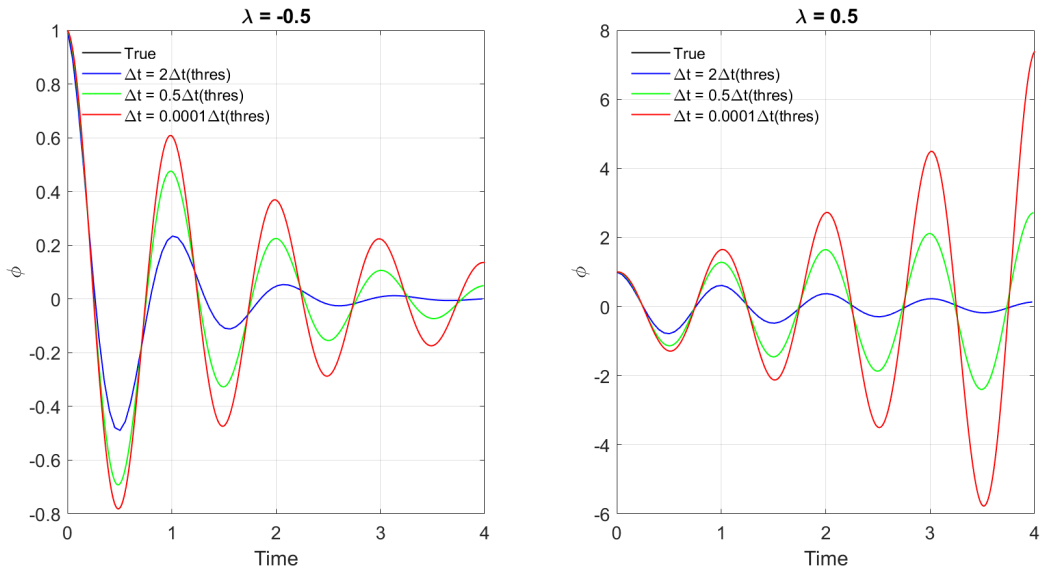
Here is the code for problems 2 - 4. The yellow highlighted sections are the additions/changes that I made in order to account for the different values of lambda & delt. I would simply comment out the values that I did not need for the solution I was employing.

```
% ========================================================================
%   This program is to examine A-stability for transport ODE
%   Numerical Scheme: forward time difference
% ========================================================================
set(0,'defaulttextfontsize',14);
set(0,'defaultaxesfontsize',14);
set(0,'DefaultAxesTickDir', 'out')
set(0,'DefaultFigureColormap',feval('jet'));

%%% define constants %%%
tau = 1;                                                    % wave period
omega = 2*pi/(tau);                                         % wave frequency
%%% define true solution parameters %%%
phi0 = 1;
deltt = 0.01;
t0  = 0; t1 = 4*tau;
tt_true = t0:deltt:t1;
%%% select different delt %%%
col = {'b','g','r'};

% FWD & BCKWD lambda values
lambda0 = [-0.5,0.5];
% Trapezoidal lambda values
lambda0 = [-0.2, 0.5];
figure(1);clf;
for ilambda = 1:numel(lambda0)
    lambda = lambda0(ilambda);
    %%% construct true solution %%%
    phitrue = phi0*exp((lambda+1i*omega)*tt_true);                  % true solution
    gamma = (lambda+1i*omega);
    delt_shred = abs(-2*lambda/(lambda^2+omega^2));

    % FWD delt values
    delt0 = [delt_shred*1.1,delt_shred*0.11,delt_shred*0.011];  % choose different delt
    % BCKWD delt values
    delt0 = [delt_shred*2,delt_shred*0.5,delt_shred*0.0001];
    % Trapezoidal delt values
    delt0 = [delt_shred*1.1,delt_shred*0.11,delt_shred*0.011];

    %%% plot true solution %%%
    figure(1);
    subplot(1,2,ilambda)
    plot(tt_true,real(phitrue),'k','linewidth',1);hold on;
    %%% solve ODE %%%
    for idtt = 1:numel(delt0)
        delt = delt0(idtt);
        tt = t0:delt:t1;

        phi = nan(size(tt));
        phi(1) = phitrue(1);
        for itt = 2:numel(tt)
            % Forward
            phi(itt) = phi(itt-1)*(1+gamma*delt);
            % Backward
            phi(itt) = phi(itt-1)/(1-gamma*delt);
            % Trapezoidal
            phi(itt) = phi(itt-1)*(1+gamma/2*delt)/(1-gamma/2*delt);
        end
        plot(tt,real(phi),col{idtt},'linewidth',1);
    end
    xlabel('Time');ylabel('\phi');title(['\lambda = ',num2str(lambda)]);grid on;
    % NOTE: I changed the legend each time to reflect the different outputs for each case
    lg = legend('True','\Deltat = 1.1\Deltat(thres)','\Deltat = 0.11\Deltat(thres)',...
        '\Deltat = 0.011\Deltat(thres)','location','northwest');
```

```
    set(lg,'box','off','fontsize',12);
end
```

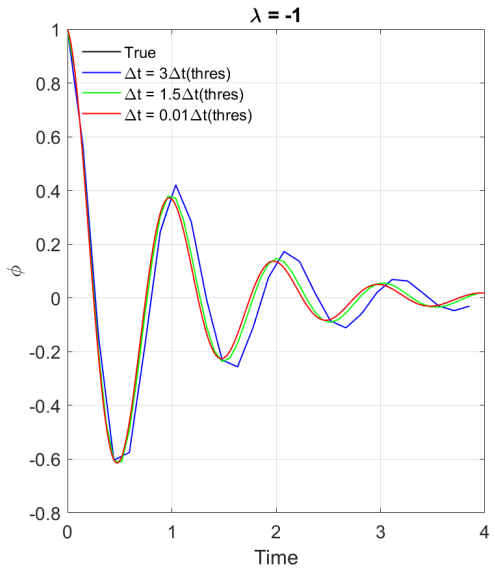
Problem 2:

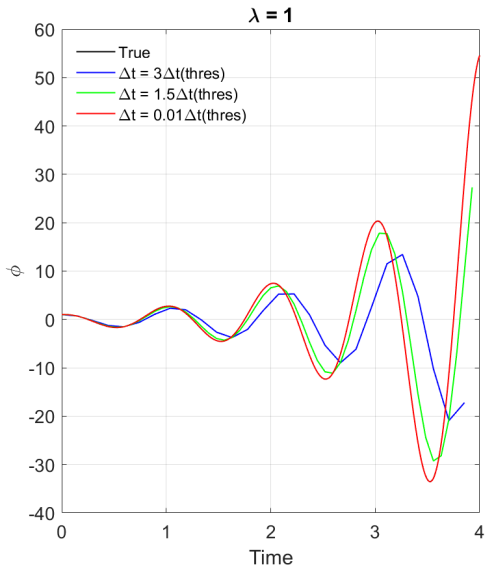
I ran the code above with the backward Euler method and the corresponding new *delt* values (2**delt_thresh*, 0.5**delt_thresh*, 0.0001**delt_thresh*). The *lambda* values remained the same for this case as in the forward Euler (-0.5, 0.5). From my observations, the smallest *delt* value, 0.0001 * *delt_thresh*, was the most accurate value of them all. The *delt* value of 2**delt_thresh* was another interesting case to look at. This interval was too large to provide convergence to the solution, and as such, it decreased over time in both cases. The accuracy of the solution increases with an decrease in *delt*, which is exactly what should happen. Below is the result (the red and black lines are almost identical, so you cannot see the black line):



Problem 3:

For the solution of problem 3, I ran the code above with the Trapezoidal numerical scheme. I adjusted the *lambda* values to be (-1.0, 1.0), and the *delt* values to be 3**delt_thresh*, 1.5**delt_thresh*, and 0.01**delt_thresh*. These values resulted in the most accurate and stable case coming from the lowest value for *delt*, which was 0.01**delt_thresh*. The least accurate case was the case for 3**delt_thresh*, which was not surprising. Accuracy of these solutions can be measured by amplitude magnitude and phase shift. The larger that *delt* became, the smaller the amplitude of oscillations became and the more the solution's phase tended to delay. These results are consistent with what should be expected of a trapezoidal scheme.

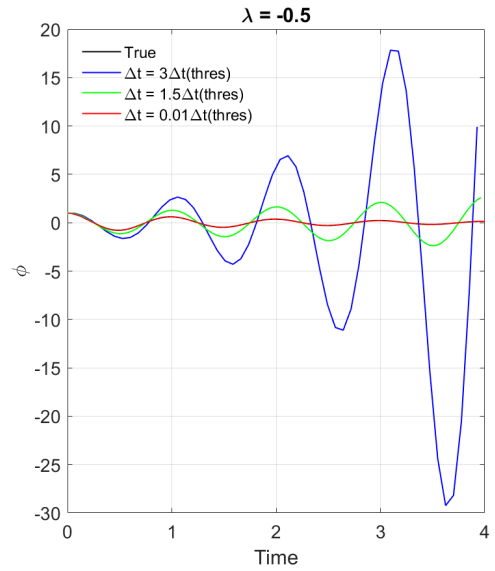


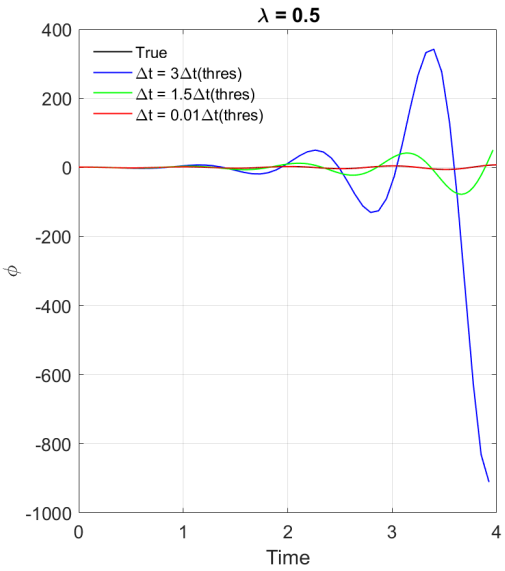


Problem 4:

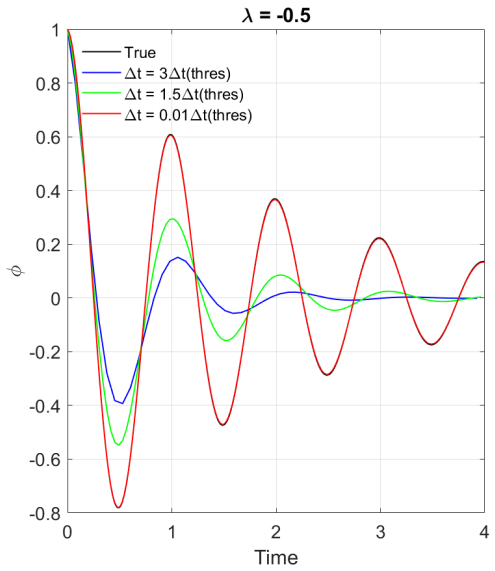
For this problem, each of the solution methods (forward-Euler, backward-Euler, and trapezoidal) were compared to each other. As such, each method was run with the same two values of *lambda*, [-0.5, 0.5] and the same *delt* steps, (3**delt_thresh,* 1.5**delt_thresh*, 0.01**delt_thresh*). Here are the results, and an analysis will follow:

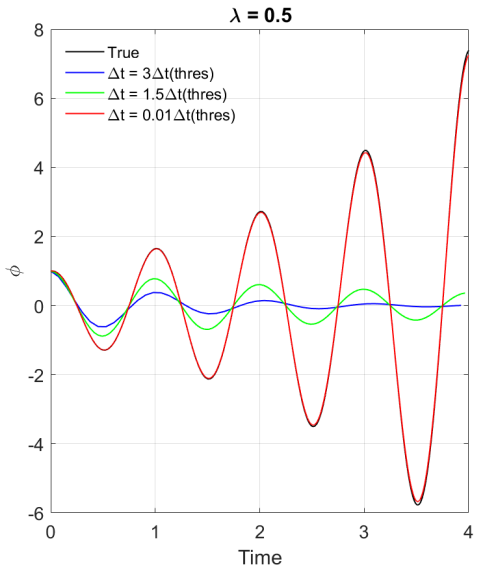
Forward:



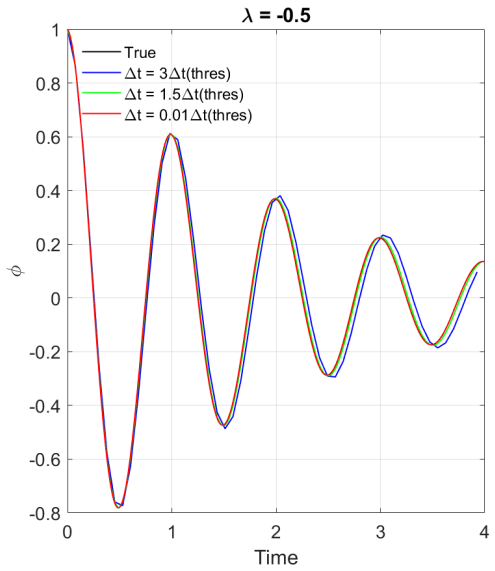


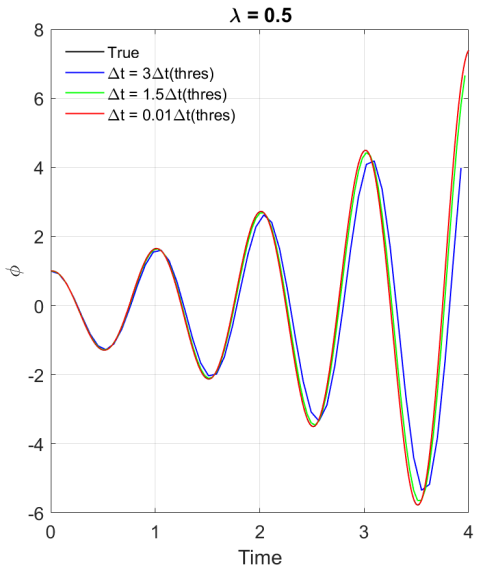
Backward:





Trapezoidal:





The criteria considered for comparing these scheme are stability, convergence, accuracy, and consistency. Looking at the graphs, the scheme with the best stability to large Δt values is the trapezoidal method. The backward-Euler method can handle a $\Delta t > 1$, but not one that is much greater than 1 without becoming unstable, and the forward-Euler method is not able to handle a Δt value greater than one without growing much more rapidly than the true solution. The most stable method is the trapezoidal method.

Looking next at convergence, problem 1 in the homework showed that the forward and backward-Euler methods had first-order convergence. The trapezoidal method, however, had a

second-order convergence, so it has less error than the other two. Therefore, the trapezoidal method has the best convergence of the three.

Examining the accuracy of these three methods entails looking at both amplitude and phase change over time. The forward method increases the amplitude of the solution over time, and it delays the phase. The backward method dampens the amplitude over time, and like the forward method it also delays the phase. The trapezoidal method maintains a constant solution amplitude over time. It too introduces a phase delay into the solution as time increases; however, its phase delay is $1/4^{th}$ that of the forward and backward schemes. Given this evaluation, the trapezoidal method is the most accurate of the three.

The final criteria for comparison is consistency of answers. Each of the three methods is has consistency in error as $\Delta t \rightarrow \infty$, but the trapezoidal method has the best consistency. It has second-order consistency whereas the Euler methods on have first-order consistency. As such, the trapezoidal method is the most consistent.

In summary, the best of these three methods, as evaluated by the criteria of stability, convergence, accuracy, and consistency, is the trapezoidal method. This is confirmed by examining the solutions to the same problem as shown in the figures above.