Project Milestone 3 – Algorithm Evaluation and Improvements

Instructions

1. Read this document carefully. You are responsible for following all instructions in this document.
2. Read the Learning Objectives at the end of the document to understand how your work will be graded.
3. Use professional language in all written responses and format all plots for technical presentation. See EPS01 and EPS02 for guidelines.
4. Apply good programming standards to all m-files.
5. Submit deliverables to Gradescope. Name your files to match the format in the table below, where *SSS_TT* is your section and team ID (e.g., 001_03 is Section 001, Team 3)

Item	Deliverables
M3 Answer Sheet	M3_*SSS_tt*.pdf
M3 Algorithm	M3_Algorithm_*SSS_tt*.m
Reference Enzyme Analysis	M3_MM_PGOX50_*SSS_tt*.m
M3 Main Function	M3_main_*SSS_tt*.m

See submission requirements on the last page of this answer sheet.

6. Complete the Assignment Header before starting the answer sheet.

Assignment Header

Section and Team ID (SSS_TT):	002_02

Team Member Name	**Purdue Career Account Login**
Connor Damato	Damato0@purdue.edu
Kush Gogia	Kgogia@purdue.edu
Jack Swingle	Jswingle@purdue.edu
Matthew Imm	Mimm@purdue.edu

Role of Each Team Member

In this section, put each team member's name who worked on this milestone. In the Detailed Description of Work, each person on the team should write their own description of how they contributed to this milestone. Be very detailed here. Then in the last column, your team should estimate the percentage of the work that each team member did on the milestone. This column needs to add up to 100%. We know that on any given milestone that this will vary, but one person in the team should not be doing significantly more than the others throughout the whole project. Use this column as a way for you to make sure your workload is balanced throughout the project.

Team Member Name	Detailed Description of Work	Percent of Work
Connor Damato	Helped debug code, coded main function, and got figures for the assignment	25
Kush Gogia	Helped with plan for M4, functions to fix.	25
Jack Swingle	Helped code the main function and fix existing functions	25
Matthew Imm	Helped answer questions, review feedback, and calculate the error	25

Part 0: M2 Feedback Review

Reflect on your M2 feedback for the purpose of improvement. Your reflection should provide a clear, useful summary of your M2 feedback and provide a clear and practical plan to address the issues. Complete table 1 below.

Table 1. Feedback summary and plan

Part A: Based on your feedback from M2, identify at least one strength and one limitation of your team's algorithm you created in M2. Consider how the feedback from M2 could lead to improvements in your work. A limitation of our algorithm is the pseudocode that went into its development. We could have included more details specifically about how our code works.
Part B: Explain how you will incorporate the M2 feedback to improve your parameter identification algorithm (do not just reword your response from Part A, include concrete actions you will take to improve). We could have included more details specifically about how our code works.

Part 1: Algorithm Development

In Milestone 2, you developed an algorithm for identifying the enzyme parameters v_{0i}, V_{max}, and K_m. Incorporate any improvements based on the feedback you received on M2 and the information in Table 1. Name the updated algorithm **M3_Algorithm_*SSS*_*tt*.m**. The algorithm function must have appropriate inputs and outputs.

Part 2: Output Comparison

The next step is to examine how well your algorithm performs using real data. NaturalCatalysts has provided data in a .csv file for a well-understood enzyme in their catalog, PGO-X50. You will use these data to examine the performance of your algorithm.

Part 2A: Understand Michaelis-Menten model for PGO-X50

Before you evaluate your algorithm, you need to understand how well the PGO-X50 data follow the Michaelis-Menten model. This information will help you understand how much error is associated with the data provided versus the parameter identification of your algorithm. Part 2A is just looking at how much error is inherent in the data, as the PGO-X50 enzyme is well-known. Therefore, in Table 2, you will find the expected parameters for PGO-X50 if the data perfectly follow the Michaelis-Menten model. You will use these reference values for the PGO-X50 enzyme following the procedures described below to determine the error in the measured data provided to you from NaturalCatalysts.

Table 2. PGO-X50 reference values

Parameter (µM/s)	PGO-X50 Reference Values
v_{01}	0.025
v_{02}	0.049
v_{03}	0.099
v_{04}	0.176
v_{05}	0.329
v_{06}	0.563
v_{07}	0.874
v_{08}	1.192
v_{09}	1.361
v_{010}	1.603
V_{max}	1.806
K_m (µM)	269.74

Use these reference values to create a user-defined function, named **M3_MM_PGOX50_*SSS*_*tt*.m**, that quantifies how well the PGO-X50 data follow the Michaelis-Menten model. This function must perform the following analyses.

1. Create a figure or set of figures to display the ten concentration curves from the given data (found in Data_PGOX50_enzyme.csv). Overlay the initial velocity, v_{0i}, tangent lines on these data. Because of the large number of data points and because the data points are sequential, you may display the measured data with a line (without using markers too). Use different styles to differentiate between the data and the tangent line. When deciding on number of figures and plots, consider what may be useful for the client to see.

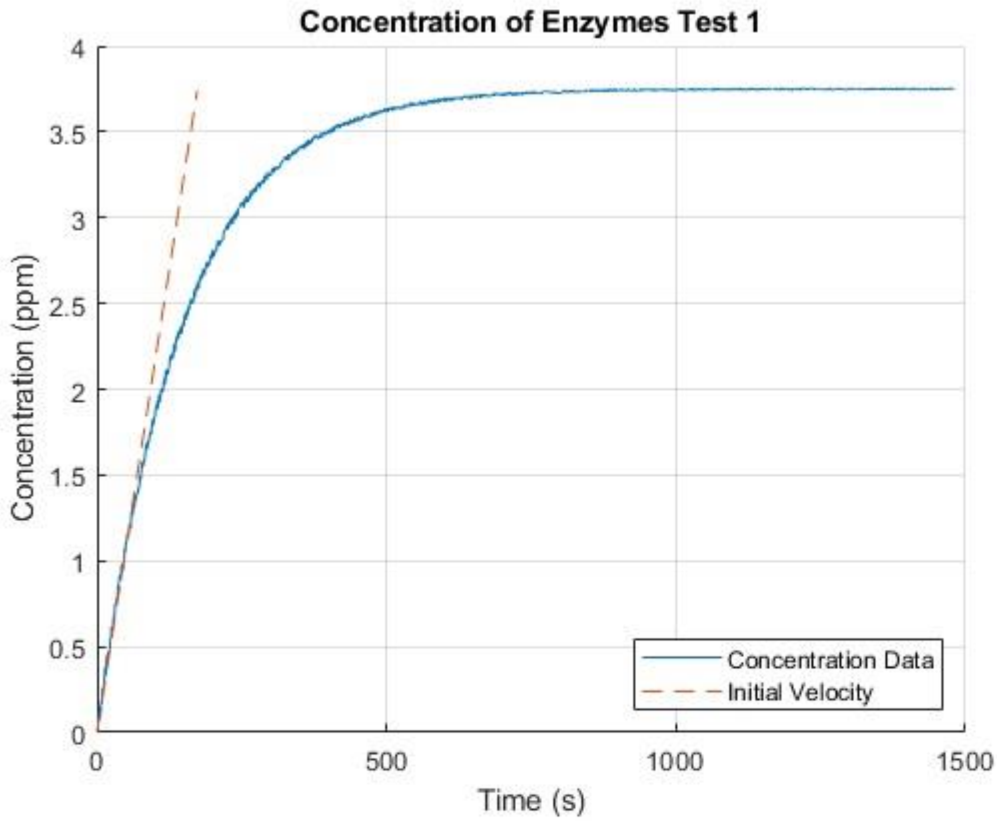
2. Create a figure that displays the initial velocities (v_{0i}) as a function of substrate concentration for the data. Plot these 10 points with data markers, since they represent the original data. Using the reference V_{max} and K_m, calculate a vector of reaction velocities using the MichaelisMenten equation (see the *ENGR 132 Project – Background and Schedule* document). Plot the model (using a solid line) along with the ten points (using markers).

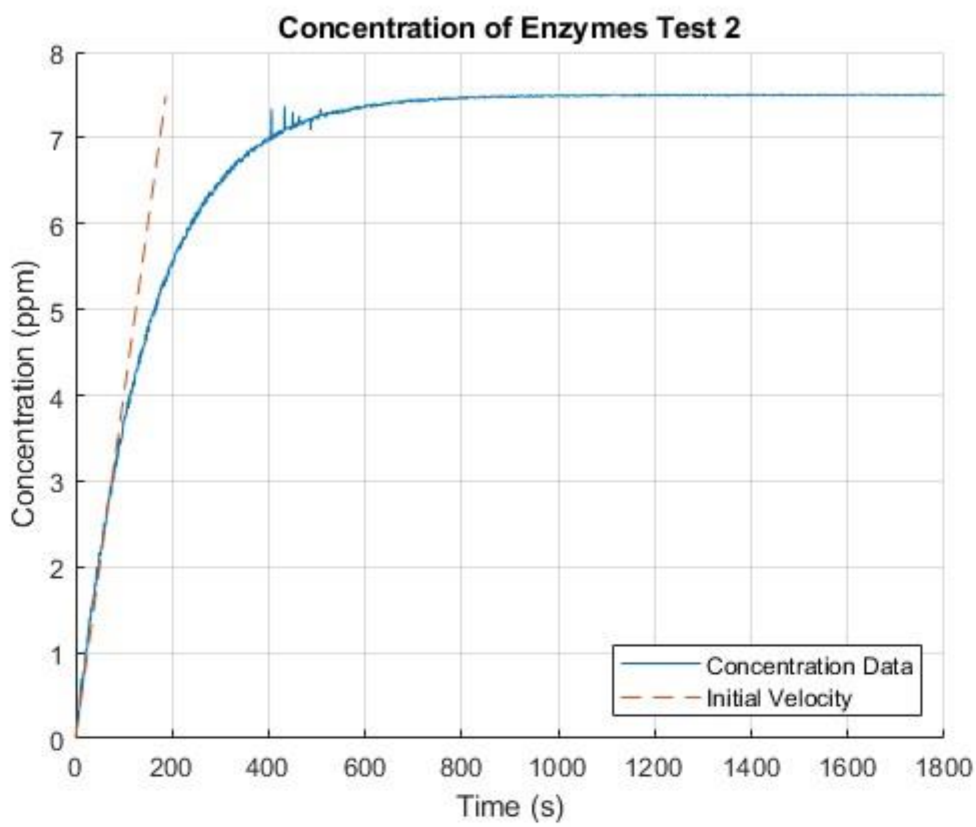
3. Estimate the error between the reference initial velocities (v_{0i}) and the reaction velocities (v_i) predicted by the Michaelis-Menten model. Use the Sum of Squared Errors (SSE) method.
4. Analyze how well the PGO-X50 data follow the Michaelis-Menten model using evidence-based rationales.

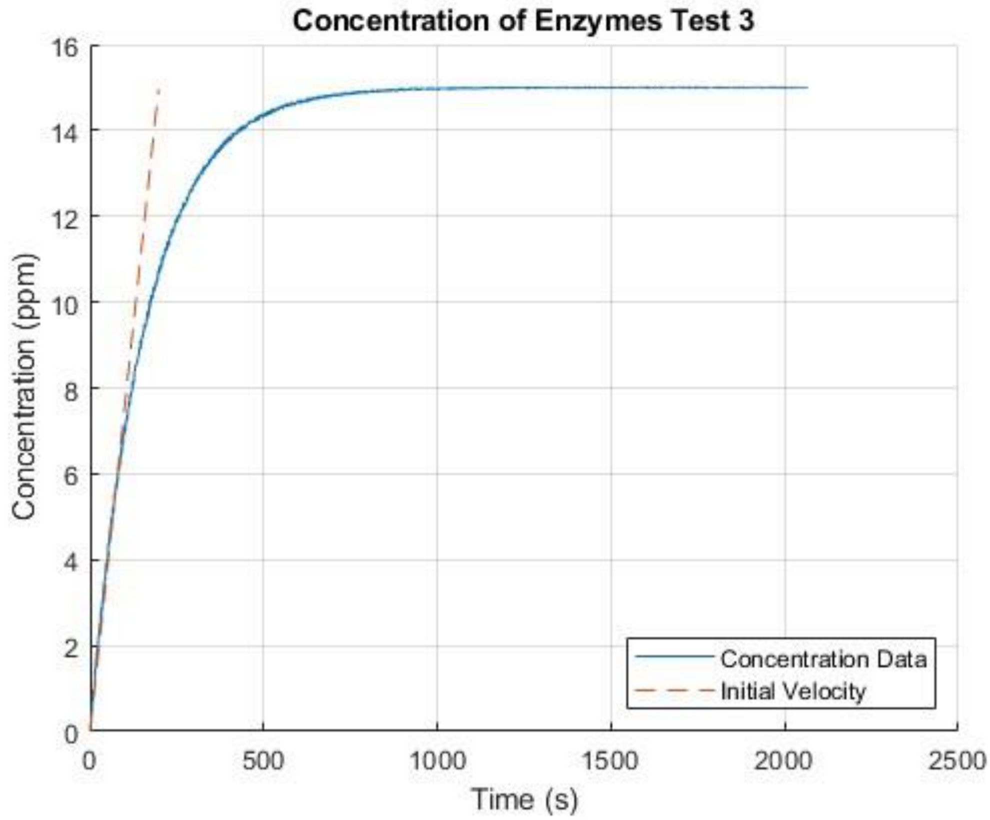
This MATLAB reference page shows how to add Greek letters and subscripts to plot displays.

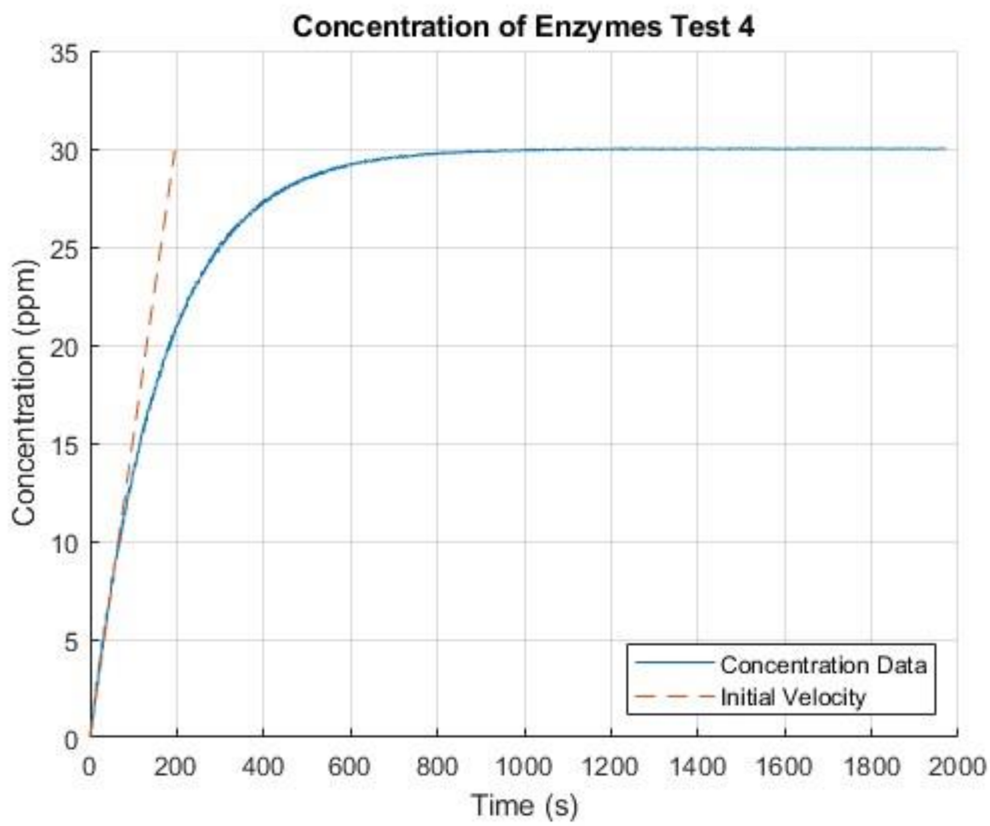
Report your figures and SSE values in Table 3. Justify your plot display choices for Steps 2A.1 and 2A.2.

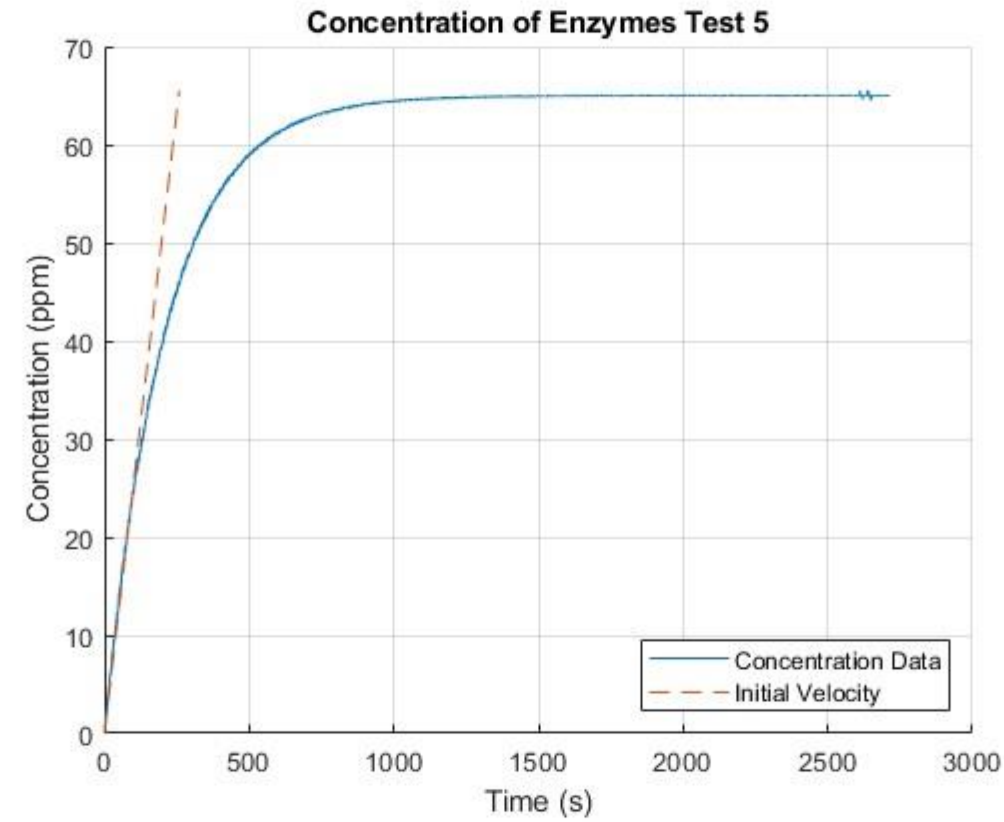
Table 3. Results for Part 2, Section A

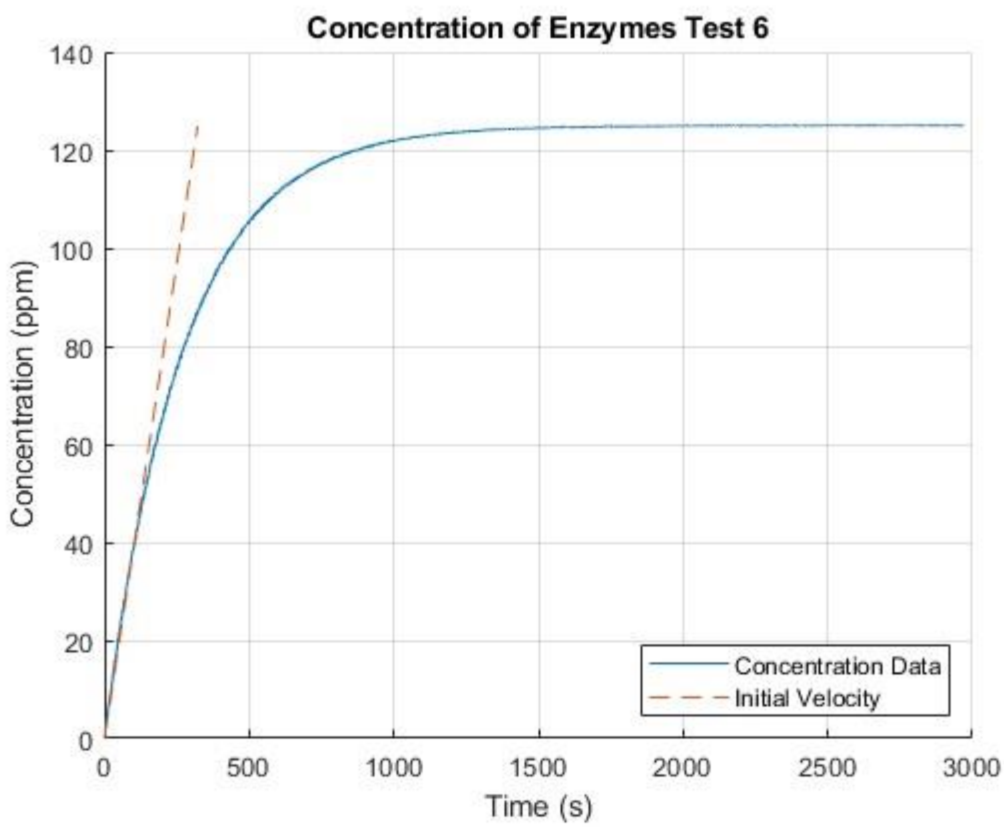

Concentration of Enzymes Test 1
Concentration (ppm)
Time (s)
0
0.5
1
1.5
2
2.5
3
3.5
4
0
500
1000
1500
Concentration Data
Initial Velocity

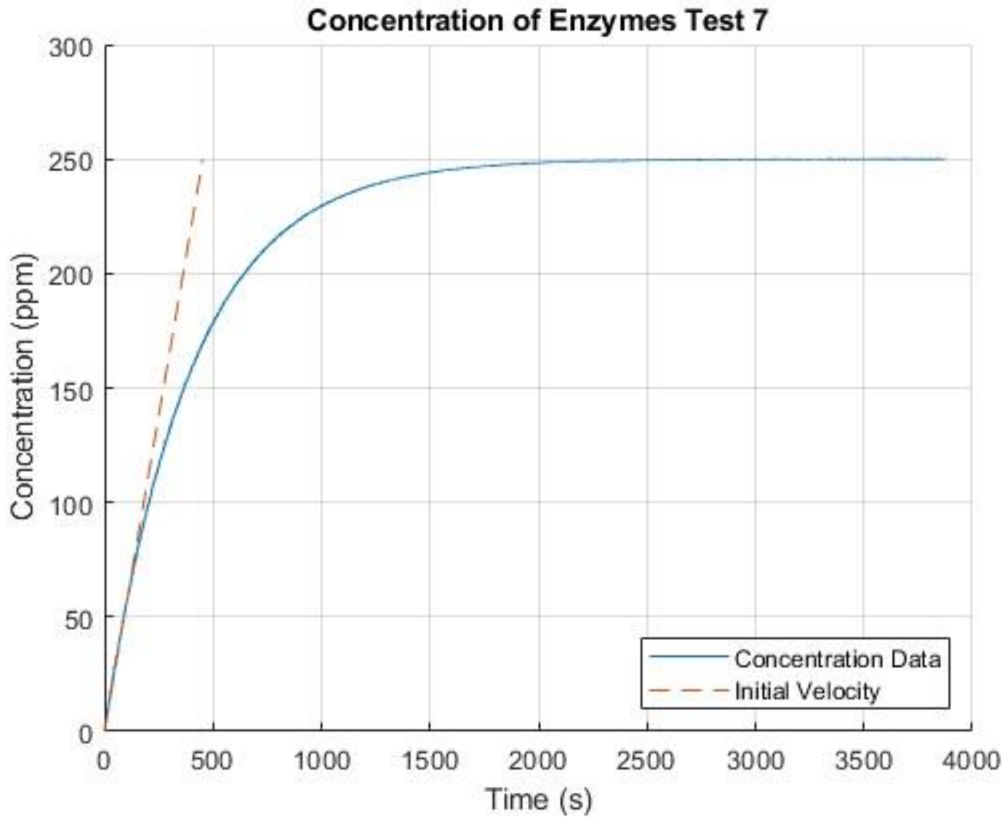

Concentration of Enzymes Test 2
Concentration (ppm)
Time (s)
0
1
2
3
4
5
6
7
8
0
200
400
600
800
1000
1200
1400
1600
1800
Concentration Data
Initial Velocity

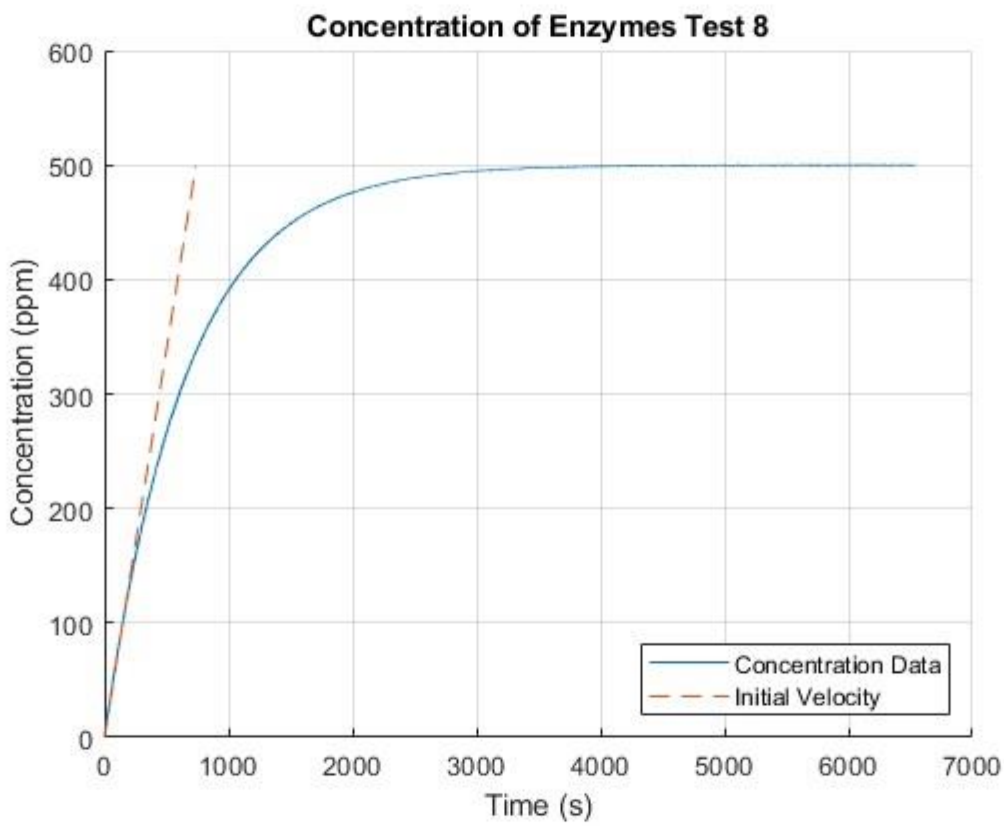

Concentration of Enzymes Test 3
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
2
4
6
8
10
12
14
16
0
500
1000
1500
2000
2500

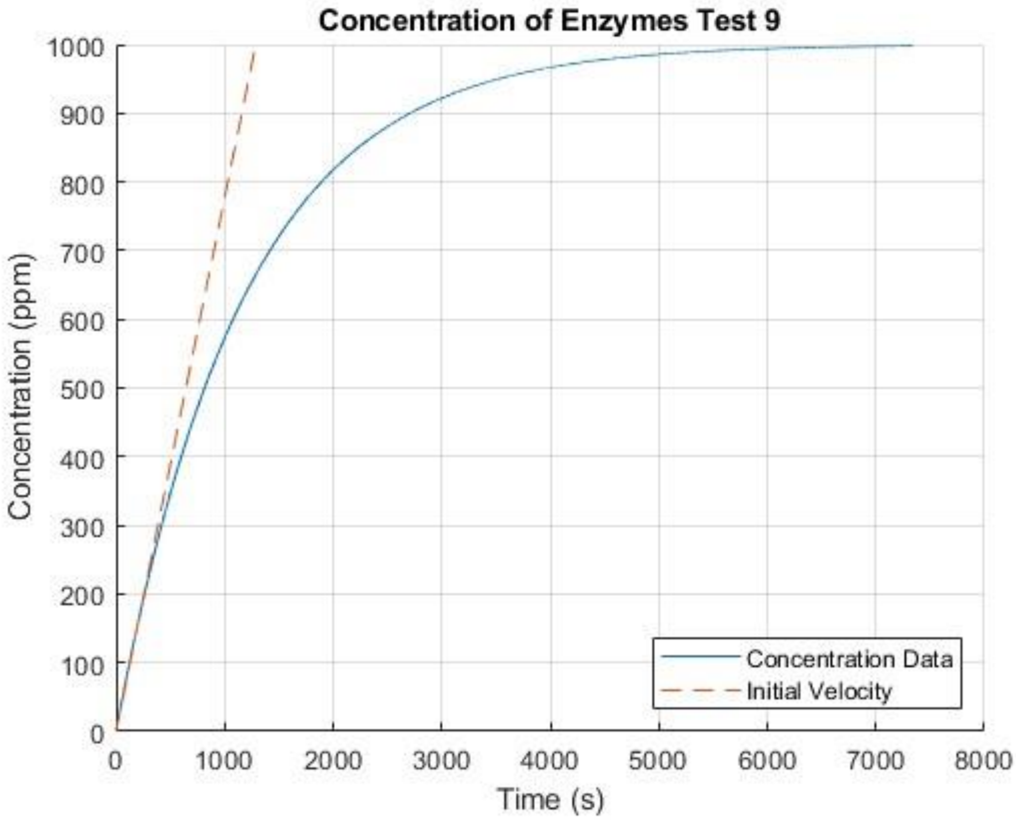

Concentration of Enzymes Test 4
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
5
10
15
20
25
30
35
0
200
400
600
800
1000
1200
1400
1600
1800
2000

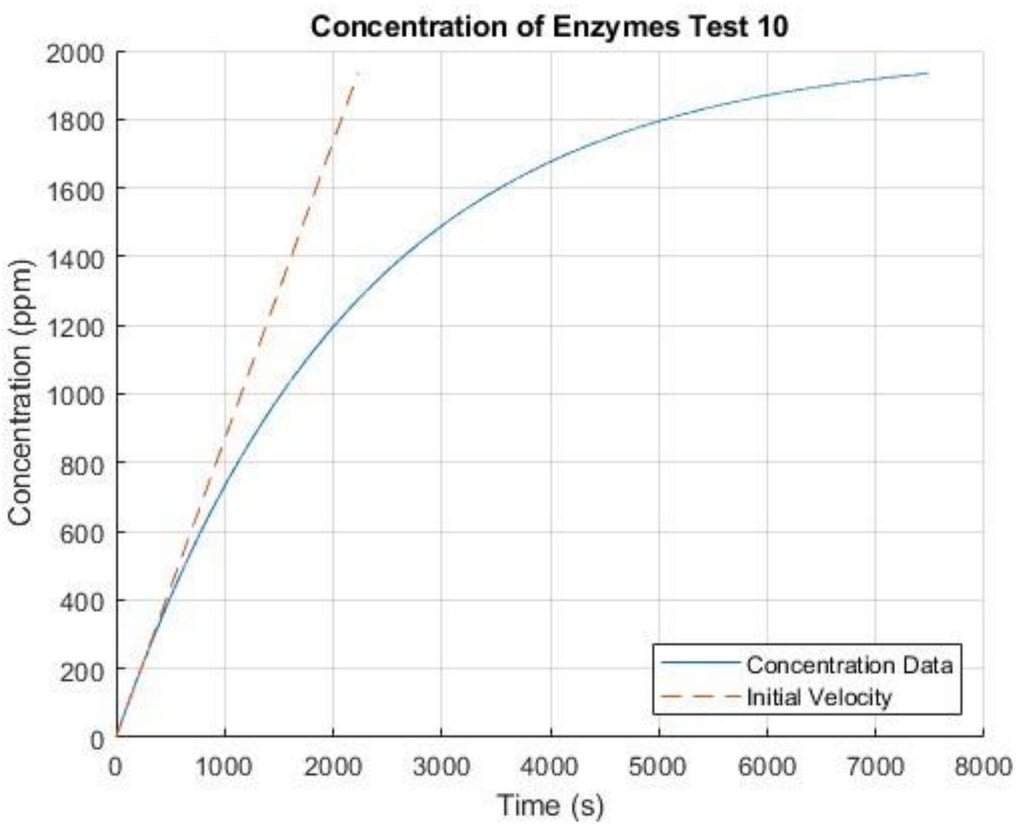

Concentration of Enzymes Test 5
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
10
20
30
40
50
60
70
0
500
1000
1500
2000
2500
3000

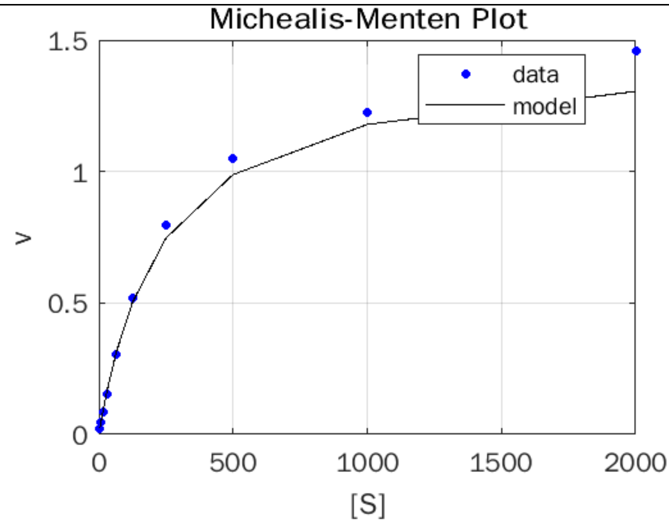

Concentration of Enzymes Test 6
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
20
40
60
80
100
120
140
0
500
1000
1500
2000
2500
3000


Concentration of Enzymes Test 7
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
50
100
150
200
250
300
0
500
1000
1500
2000
2500
3000
3500
4000


Concentration of Enzymes Test 8
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity
0
100
200
300
400
500
600
0
1000
2000
3000
4000
5000
6000
7000


Concentration of Enzymes Test 9
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity


Concentration of Enzymes Test 10
Concentration (ppm)
Time (s)
Concentration Data
Initial Velocity



Justify your plot display choices for Steps 1 and 2: Using multiple plots allowed for easier viewing and it is still possible to compare the data side-by-side.
SSE for data: 0.033
Analysis of SSE: The SSE is best when as low as possible, so this value seems pretty reasonable. However, the values themselves are pretty small, so this SSE may not represent how relatively low high error the data are.

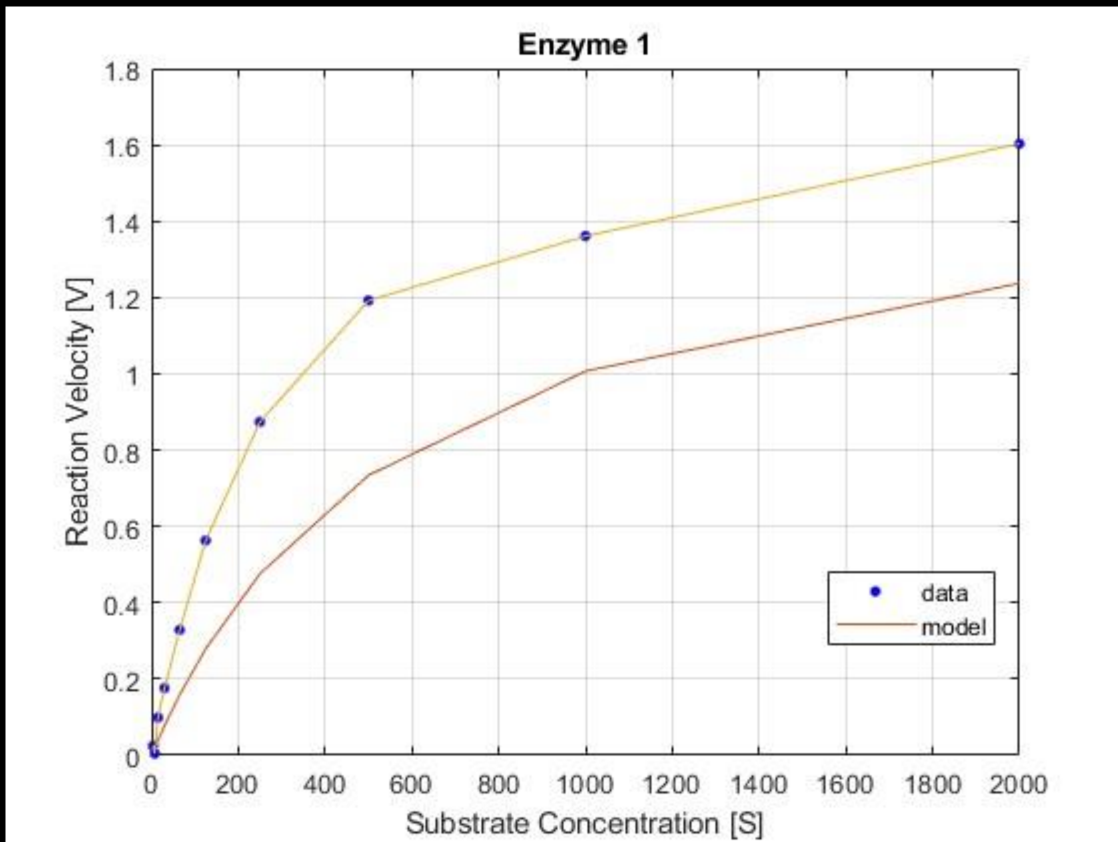
Part 2B: Main function using your algorithms

Now you will update your main function that will call your algorithm UDF (from Part 1) so that you can examine how well your algorithm identifies the parameter values from the Michaelis-Menten model, as compared to the reference values given to you in Part 2A. Using the data for PGO-X50, determine the enzyme parameters using your algorithm. Name the main function **M3_main_*SSS_tt*.m**. Remember, a main function has no input arguments and no output arguments. The function must perform the following tasks:

1. Load the PGO-X50 data.
2. Call your parameter identification algorithm to determine the parameters for the PGO-X50 data.
3. Plot the data and the Michaelis-Menten model using the steps 1 and 2 described in Part 2A (above). Remember to consider readability and usefulness to the client.
4. Determine the SSE for the PGO-X50 data.

Report your figures in Table 4.

Table 4. Figure displays for your Algorithm



Figures for Algorithm

hm

[redacted]

Evaluate your algorithm for finding $v_0's$: Complete Table 5a using the v_0 parameters from your algorithm and the model using the reference parameters v_{max} and K_m. Calculate the percent error and SSE results.

Table 5a. Algorithm comparison of v'_0s using data results

Parameter (µM/s)	PGO-X50 Reference Values	M2_Algorithm	Percent Error
v_{01}	0.025	0.0223	10.8
v_{02}	0.049	0.0476	2.86
v_{03}	0.099	0.0870	12.12
v_{04}	0.176	0.1555	11.65
v_{05}	0.329	0.3030	7.9
v_{06}	0.563	0.5179	8.01
v_{07}	0.874	0.7967	8.84
v_{08}	1.192	1.0501	11.9

$v0_9$	1.361	1.2249	10
$v0_{10}$	1.603	1.4607	8.88
V_{max}	1.806	1.4607	19.1
K_m (µM)	269.74	238.93	11.4
SSE	(V0) 0.06816		

Evaluate your algorithm for finding v_{max} and K_m: Complete Table 5b using the reference v_0 parameters and the model using the parameters v_{max} and K_m from your algorithm. Calculate the percent error and SSE results.

Table 5b. Algorithm comparison of v_{max} and K_m using data results (Updated algorithm)

Parameter (µM/s)	PGO-X50 Reference Values	M2_Algorithm	Percent Error
$v0_1$	0.025	0.0223	10.8
$v0_2$	0.049	0.0476	2.86
$v0_3$	0.099	0.0870	12.12
$v0_4$	0.176	0.1555	11.65
$v0_5$	0.329	0.3030	7.9
$v0_6$	0.563	0.5179	8.01
$v0_7$	0.874	0.7967	8.84
$v0_8$	1.192	1.0501	11.9
$v0_9$	1.361	1.2249	10
$v0_{10}$	1.603	1.4607	8.88
V_{max}	1.806	1.4607	19.1
K_m (µM)	269.74	238.93	11.4
SSE	(Vmax) 0.119232	(Km) 949.2561	

Part 3: Observations and Improvements

Based upon your observations of your plots and the SSE results for your model, suggest at least two ways you believe your algorithm could be improved. Briefly explain each suggestion below using evidence-based rationales. You do not need to code these changes; at this point, simply describe changes you think might be useful.

Be sure to:

- explain which parameter(s) your improvement will target,

- explain the improvement with a level of detail that can be understood by others (provide sketches or flowcharts as necessary to clarify your improvement),
- describe the performance metrics you will use to determine whether your proposed improvement really does improve your solution, and
- provide evidence-based rationales for each proposed improvement and the metrics selected. Your rationales should answer the questions:
  - What is your evidence that this improvement is necessary? ○ Why is this method for making the improvement a good idea - what is your evidence?
  - Why is this metric a good idea - what is your evidence?

Discuss improvements that resolve issues that you see in your current algorithm in the Table 6 below. Cite any external references in Table 8.

If, based on your PGO-X50 analysis, you feel that you do not need to make changes to your algorithm, indicate your rationale in Table 7. Cite any external references in Table 8. You should start thinking ahead to the data sets you will work with in M4 (data from ten concentration tests plus ten duplicate tests for 5 enzymes [i.e., the data from earlier milestones]).

Table 6. Algorithm improvements (add improvement blocks as needed)

Improvement 1. Parameter(s) Targeted: V0
Description Our V0 values were continuously underestimated. This is because our algorithm found the best r^2 line for the entire data instead of trying to get the smallest accurate domain to get the line.
Metrics to Determine Improvement Percent error of Km- We can look at this singular number to tell quickly if a solution works or not. Graph- We can view the graph to see what kind of general mistakes are made. We can see if it is an underestimate or overestimate and quickly view any outliers to the data. We were able to spot this error though our calculation of the percent error against the actual values. This wasn't seen as well through the graphs, since the data is off by such a small margin, however it is more clear in the numbers.
Rationale for Improvement and Metrics This improvement is necessary because if we continuously underestimate the values, we will also underestimate our Km values. This means if our company is comparing it to its competitors Km values, ours might be low.
Improvement 2. Parameter(s) Targeted: Vm

Description
The Vmax value was extremely underestimated, most likely because we were using the Vmax as the maximum value of the calculated v0s instead of calculating it from the linearization.
Metrics to Determine Improvement
Percent Error of Vmax – decreasing the percent error between the theoretical and experimental values of Vmax will help determine the fact that any changes made in the algorithm are good Plot visual match: We can see if the Vmax is getting correct if the Michealis-Menten model more accurately represents the data given. In this case, the Michealis-Mentin model does not fit the data well, and it shows that it is off by quite a large margin, meaning our vmax values are most likely off.
Rationale for Improvement and Metrics
The improvement is necessary because incorrect Vmax values can reduce the trust in the model, which is a huge chunk of what the client needs.

Table 7. Justification for not needing improvements to algorithm

--

Table 8. References used in evidence-based rationales

--

How to Submit

1. Save this answer sheet as a PDF named **M3_*SSS*_*tt*.pdf** where ***SSS*** is your section number (e.g., 001 for section 001) and ***tt*** is your team number (e.g., 07 for team 7).
2. Select one person to submit the deliverables for the team. That person should
   a. Log into Gradescope and submit all deliverables to the **M3** assignment.
      i. M3_*SSS*_*tt*.pdf
      ii. M3_Algorithm_*SSS*_*tt*.m
      iii. M3_MM_PGOX50_*SSS*_*tt*.m iv. M3_main_*SSS*_*tt*.m
   b. Select all team members for the group assignment and submit.
   c. Double-check that all team members are assigned to the submission.
3. Each team member should confirm that they are part of the submission.
4. After submission, distribute the submitted files to all team members. *Ensure all members of the team have copies of the submitted files.*

Learning Objectives

Teamwork (TW)
Contribute to team products and discussions
TW02. Document all contributions to the team performance with evidence that these contributions are significant.

Process Awareness (PA)
Reflect on both personal and team's problem solving/design approach and process for the purpose of continuous improvement.
PA01. Identify strengths in the approach used.
PA02. Identify limitations in the approach used.
PA03. Identify potential behaviors to improve approach in future problem solving/design projects.

Evidence-Based Decision Making (EB)
Use evidence to develop and optimize solution. Evaluate solutions, test and optimize chosen solution based on evidence.
EB01. Test prototypes and analyze results to inform comparison of alternative solutions.
EB03. Clearly articulate reasons for answers with explicit reference to data to justify decisions or to evaluate alternative solutions.
EB05. Present findings from iterative testing or optimization efforts used to further improve aspect or performance of a solution.
EB06. Clearly articulate reasons for answers when making decisions or evaluating alternative solutions.

Solution Quality (SQ)
Design final solution to be of high technical quality. Design final solution to meet client and user needs. SQ01. Use accurate, scientific, mathematical, and/or technical concepts, units, and/or data in solutions.

Data Visualization and Analysis (DV)
Visually represent data and derive meaningful information from data. Demonstrate ability to accurately describe data sets through foundational descriptive statistics and then perform simple inferential statistics through understanding probability concepts and linear regression. DV03. Justify graphical representation based on data characteristics.

Information Literacy (IL)
Seek, find, use and document appropriate and trustworthy information sources. IL03. Support all claims made with that is either generated or found

Engineering Professional Skills
PC05. Fully address all parts of assignment by following instructions and completing all work.

EPS01. Use professional written and oral communication.

EPS02. Format plots for technical presentation.

Programming
MAT01. Develop code that follows good programming standards

MAT08. Debug scripts and functions to ensure programs execute properly, perform all required tasks, and produce expected results.