Project Milestone 4 – Algorithm Refinement and Final Deliverable

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. Good programming standards apply 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
M4 Answer Sheet	M4_AnswerSheet_*SSS_TT*.pdf
M4 Algorithm	M4_Algorithm_*SSS_TT*.m
M4 Main Function	M4_Main_*SSS_TT*.m
Technical Brief	M4_TechnicalBrief_*SSS_TT*.pdf

   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):	LC1_14

Team Member Name	**Purdue Career Account Login**
Jun Shern Lim	lim321
Kevin Crowley	crowlek
Shravan Ranganathan	rangana2
Tyler Maslak	tmaslak

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
Jun Shern Lim	Algorithm refinement: MATLAB code refinements, Table 2, Technical Brief Introduction	25
Kevin Crowley	M3 Feedback Review, Technical Brief: Results and References	25
Shravan Ranganathan	Resume Insert, Algorithm Refinement Implementations	25
Tyler Maslak	Technical Brief: Interpretations, Appendix and Figures	25

Part 0: M3 Feedback Review

Reflect on your M3 feedback for the purpose of improvement. Your reflection should provide a clear, useful summary of your M3 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 M3, identify at least one strength and one limitation of your team's work in M3. Consider how the feedback you received on M3 could lead to improvements in your work. One strength of our algorithm is its versatility. We used a selection structure that detects input to ensure that our algorithm can accept either just original enzyme data or original and duplicate enzyme data and still perform as intended. One weakness of our algorithm is that the calculated Km and Vmax values still defer from the reference Km and Vmax value by quite a bit. This could be because we used the Eadie-Hofstee plot, which is known for amplifying experimental error.
Part B: Explain how you will incorporate the M3 feedback to improve your parameter identification (do not just reword your response from Part A; include concrete actions you will take). Although we did not receive feedback on our parameter identification in M3, we will improve our v0, Km and Vmax values by averaging the slope of the first 39 points instead of 50. We will also use the mode average of the slopes instead of the mean average to calculate our parameters. We did extensive experimentation and found that this approach gave us Km and Vmax values that were closer to their reference values. We will explain this further in Table 2 below. Additionally, we will improve the accuracy of our Vmax and Km values by using the Hanes-Woolf plot instead of the Eadie-Hofstee plot to calculate these parameters. The Hanes-Woolf plot is known to be more accurate Marasović et al. (2017), and our experimentations have proven so as well. We will explain this further in Table 2 below.

Part 1: Algorithm Improvements Plan

Respond to each of the prompts below in the space provided. Your goal is to introduce the **two improvements** to your M3 algorithm. Use your ideas from Part 3 of M3 to help formulate ideas. Briefly describe, in words (not code), the nature of the improvements you will implement in your MATLAB code. Provide a brief, but thoughtful, description of your refinement, using evidence-based rationales for why the refinement is necessary and should improve your solution. Read the rest of this document carefully ***before*** you begin your work on this milestone. Once you are ready to begin Part 1, put your refinements and your rationale in Table 2.

Table 2. Algorithm refinement plans

Refinement 1
Parameter(s) Targeted: v0, Km and Vmax Description In M3 we found the average (mean) slope values of the first 50 points to determine the v0 at each substrate concentration. To improve on that, we will use the mode instead of the mean to determine the average of the slopes. Additionally, we will take the average slope of the first 39 points instead of the first 50 points.
Rationale for Refinement After experimentation, we found that using the mode produces v0 values that output a lower SSE and more accurate Vmax and Km values when compared to the reference values. This is most likely because the mode takes the most frequently occurring slope as the average, removing the possibility of extreme values affecting v0. We decided to use 39 points instead of 50 because the data at 39 points would be more linear than at 50 points. Additionally, using 39 points produced a Km value that was significantly closer to its reference value.
Refinement 2
Parameter(s) Targeted: Km and Vmax Description Instead of using the Eadie-Hofstee plot, we will use the Hanes-Woolf plot to linearize the enzyme kinematics data and determine Vmax and Km. $\boxed{\frac{s}{v} = \frac{K_m}{V_{\max}} + \frac{1}{V_{\max}} s,}$ *Figure 8: Hanes-Woolf plot* (Marasović et al., 2017) To linearize data according to the Hanes-Woolf plot, we will need to plot (substrate concentration [S]/reaction velocity) against substrate concentration[S]. The Vmax will be 1/slope and the Km would be the y-intercept * Vmax.
Rationale for Refinement According to the journal article by Marasović et al. (2017), the Hanes-Woolf plot is the more accurate than both the Lineweaver-Burk plot and Eadie-Hofstee plot. Therefore, we believe we will get a more accurate Vmax and Km value.

Part 2: Algorithm Refinements Implementation

Before you make any changes to your code, resave your M3 code files as

- M4_Algorithm_*SSS_TT*.m
- M4_Main_*SSS_TT*.m

Implement improvements in M4_Algorithm_*SSS_TT*.m. **Clearly comment where you made improvements within the code, using the text 'Improvement 1' or 'Improvement 2' and a concise, meaningful description of the change for each improvement.**

Do not delete any code as you implement the improvements: comment out unnecessary code and comment on the change. Clearly indicate where new code is added with the commenting described above.

Evaluate the improvements in your algorithm by using the data for the reference enzyme PGO-X50 from M3. Compare the parameters identified for the PGO-X50 data using the algorithm you submitted in M3 and your refined algorithm for M4. This step ensures that you can compare the error of your algorithm to a known error in the data. Report your results in Table 3. Use appropriate decimal places.

Table 3. Algorithm refinement comparison

Parameter (μM/s)	PGO-X50 Reference Values	M3_Algorithm	M4_Algorithm
v_{0_1}	0.025	0.0264	0.0254
v_{0_2}	0.049	0.0512	0.0491
v_{0_3}	0.099	0.1020	0.0985
v_{0_4}	0.176	0.1785	0.1729
v_{0_5}	0.329	0.3483	0.3384
v_{0_6}	0.563	0.5494	0.5366
v_{0_7}	0.874	0.8559	0.8395
v_{0_8}	1.192	1.1919	1.1754
v_{0_9}	1.361	1.3439	1.3338
$v_{0_{10}}$	1.603	1.5742	1.5667
V_{max}	1.806	1.7154	1.7609
K_m (μM)	269.74	246.6724	269.6739
SSE $(\mu M/s)^2$	0.0048	0.0061	0.0046

Next, use your M4 algorithm to analyze the full 100 enzyme test data sets and obtain the parameters V_{max} and K_m. Here you will run your M3 algorithm and your updated M4 algorithm on the full data set. You may need to make adjustments to both algorithms to account for the replicate data sets and 5 enzymes. In Table 4, record your results from both your M3 and M4 algorithm. Use appropriate decimal places.

Table 4. M3 and M4 algorithm comparison of experimental data parameters

Enzyme	M3 Algorithm			M4 Algorithm		
	Enzyme Parameters		SSE $(\mu M/s)^2$	Enzyme Parameters		SSE $(\mu M/s)^2$
	V_{max} (μM/s)	K_m (μM)		V_{max} (μM/s)	K_m (μM)	
NextGen-A	0.9020	147.8701	0.0016	0.9290	165.0574	0.0017
NextGen-B	0.8039	328.9893	0.0009	0.8238	347.2961	0.0005
NextGen-C	1.1536	181.6380	0.0003	1.1673	192.6289	0.0001
NextGen-D	1.4853	280.6591	0.0011	1.4639	280.6448	0.0009
NextGen-E	1.5648	158.7816	0.0004	1.5567	169.7266	0.0003

In Table 5, include any references you used throughout this answer sheet for Parts 0-2. Use APA format. Make sure there is an in-text citation for all references listed and vice versa.

Table 5. References used in Parts 0-2 (if any)

Marasović, M., Marasović, T., & Miloš, M. (2017, March 5). *Robust Nonlinear Regression in Enzyme Kinetic Parameters Estimation*. Journal of Chemistry. https://www.hindawi.com/journals/jchem/2017/6560983/.

Part 3: Technical Brief

Consult the M4 memo from NaturalCatalysts, Inc. for the details concerning your technical brief. Use the provided template M4_TechnicalBrief_template.docx to respond to the memo. You may find the original introduction memo and the project background documents helpful when composing your technical brief.

Part 4: Resumé Insert

In response to the opportunity presented in the NaturalCatalysts memo, create an insert for your resumé by completing the following on this answer sheet:

Guidance:

Summarizing your ENGR 132 project for your resumé

Choose a header and specific language to describe your project. Possible Headers for Engineering 132 Project Descriptions include: Engineering Projects, Design Projects, Related Experience, Engineering Experience. The specific language should be "action" oriented and highlight both the project and your contributions to it. Your project title should be something that describes this project.

Template:

DESIGN PROJECT

Enzyme Analyzation, *Purdue University* *Spring 2021*

- Collaboratively developed computational models to analyze provided catalyst data.
- Presented findings in an efficient and appealing manner.
- Produced a recommendation to our client based on the findings.

Example:

DESIGN PROJECTS

Autonomous Lawn Mower, *Purdue University* *Spring 2020*

- Improved sensor technology resulting in increased safety and reduced cost
- Developed MATLAB code to optimize sensor performance and to perform constraint analysis
- Constructed and tested a functional prototype that surpassed industry standards

Things to keep in mind:

- Headers should stand out (Bold/Underlined/Larger Font and/or CAPS).
- Do not use "Engineering 132" Project as the project title. Prospective employers will not know what that title means. Give the project a descriptive name.
- Differentiate between project title and location using style change or location variance.
- Separate the location and the date of the project. Placing the date on the right side of the page is common, but not required.
- Your 3-5 bulleted statements should all maintain the *same tense* (past if previously completed, or present if currently working on).
- Begin each bullet with a different power verb.
- For these 3-5 statements, try to answer the questions "What did you do?", "How did you do it?", and "What was your result?"

Resumé Text: In the space below, write a **summary of your project suitable for inclusion on your resumé**. Be sure to use the guidelines above regarding formatting and language. A resumé typically includes 3-5 bullet items describing a project. The stems for your bullet points should be power verbs that convey what you did on the project (i.e., implemented, led, developed, analyzed, etc.). Use your individual versions from the video activity to create a team version here.

Design Project

NaturalCatalyst Enzyme Data Analysis, *Purdue University* Spring 2021

- Developed a MATLAB model to analyze a large enzyme dataset.
- Conducted error analysis and iteration to improve the efficacy of our model.
- Developed a technical brief to present our recommendations to the client.

Finally, you should each add this insert or your individual one from the video activity into your own resumé.

How to Submit

1. Save this answer sheet as a PDF named **M4_AnswerSheet_*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. Save your technical brief document as a PDF named **M4_TechnicalBrief_*SSS_TT*.pdf**.
3. Select one person to submit the deliverables for the team. That person should
    a. Log into Gradescope and submit all deliverables to the **M4** assignment.
        i. M4_AnswerSheet_*SSS_TT*.pdf
        ii. M4_Algorithm_*SSS_TT*.m
        iii. M4_Main_*SSS_TT*.m
        iv. M4_TechnicalBrief_*SSS_TT*.pdf
    b. Select all team members for the group assignment and submit.
    c. Double-check that all team members are assigned to the submission.
4. Each team member should confirm that they are part of the submission.
5. 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.

Idea Fluency (IF)
Generate ideas fluently. Take risks when necessary.
IF03. Generate testable prototypes (including process steps) for a set of potential solutions.

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.

Information Literacy (IL)
Seek, find, use and document appropriate and trustworthy information sources.
IL04. Include citations within the text (in-text citations) that show how the references at the end of the text are used as evidence to support decisions.
IL05. Format reference list of used sources that is traceable to original sources (APA or MLA are recommended)

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.

MAT05. Create and use MATLAB scripts and functions.
MAT08. Debug scripts and functions to ensure programs execute properly, perform all required tasks, and produce expected results.