Project Milestone 1 – Parameter Identification Brainstorming

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. See EPS01 for guidelines.
4. Submit deliverables to Gradescope for grading. 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	Type	Deliverables
M1 Answer Sheet	Team	M1_AnswerSheet_*SSS_TT*.pdf

   See submission requirements on the last page of this answer sheet.
5. 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**
Jack Swingle	jswingle
Matthew Imm	mimm
Kush Gogia	kgogia
Connor Damato	damato0

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 this milestone. This column needs to add up to 100%. We know this will vary on any given milestone, 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**
Jack Swingle	Involved in the brainstorming and typing of each of the tables for the problem statement and the approaches to find variables. Also helped with the ways to deal with noise.	25

Matthew Imm	Involved in the brainstorming and typing of each of the tables for the problem statement and the approaches to find variables. Also helped with the ways to deal with noise.	25
Kush Gogia	Involved in the brainstorming and typing of each of the tables for the problem statement and the approaches to find variables. Also helped with the ways to deal with noise.	25
Connor Damato	Involved in the brainstorming and typing of each of the tables for the problem statement and the approaches to find variables. Also helped with the ways to deal with noise.	25

M1 Context

Parameter identification is about determining model parameters from available data. This milestone requires you to think carefully about different potential parameter identification approaches, the process of using/coding those approaches, and the consequences of those choices.

Your assignment is to develop several ideas for fully automated identification of parameters from first-order data of enzyme reactions. You should use evidence-based rationales when evaluating your ideas. In this context, "evidence-based" means you should refer to the graph of the data provided - where the graph is a proxy for actual data, your knowledge of MATLAB, and other trusted sources when evaluating your ideas.

Note: When external sources are used, each must be properly cited with (1) an in-text citation where referenced in the body of the text and (2) a full citation in the appropriate *References* section for each part of this milestone document. Use APA 6th or 7th style.

Part 1: Team Problem Statement

In M0, each team member created an individual problem statement for the project. In Table 1, combine your problem statements into a team problem statement. Remember, a problem statement is a clear, concise, and complete description of a problem to be solved. A problem statement includes information on the need and the user (or a key stakeholder) without reference to a specific solution.

A problem statement is **your interpretation** of the clients' needs or wants.
A good problem statement (PS01 & PS02):

1. Clearly refers to a **client**
2. Clearly states the **need, problem, or the focus** of the project
3. Explains **why** this need/problem/focus is **important** to solve

followed by a list of specifications (PS03)

1. **Criteria** and **reason** for each criteria (needs/wants)
2. **Constraints** and **reason** for each constraint (required/must-haves)

Table 1. Team Problem Statement

Team Problem Statement
Commercial enzymes are used in food, beverages, detergents, pulp, textile, and many other products. One of the leading engineering companies that manufactures these enzymes is NaturalCatalysts Incorporated. Our team was hired by Natural Catalysts Inc. to analyze kinetic enzyme data. NaturalCatalysts has provided us with 100 tests, 2 trials of 10 kinetic-enzyme tests for each of the 5 enzymes. We are assigned to analyze the data in order to create sales literature for Q&H of Greater Lafayette. Along with our analysis, we need to create a graphic to represent our findings, complete with error analysis and a recommendation of how Natural Catalysts can represent their enzymes to their customers. Criteria: • The solution must accurately display data, with understandable format • The solution must show trends in the enzyme test data • The solution must analyze the error of the trend models in predicting trends • A formal presentation must accompany the solution with recommendations about the test data and how it can accurately and ethically be represented to customers Constraints: • Our project must be done by the end of the semester • We have only been given 100 enzyme tests to analyze

Part 2: Brainstorm Coding Processes for Parameter Identification

Section A: Brainstorm Processes Using Clean Data

The “clean” data used in the in-class activity to identify parameters visually were derived using real experimental data and then applying processes to change the data to fit a nice, clean curve. You will use this clean data as a proxy for actual experimental data for this part of the assignment. Use ideas from your in-class activity to brainstorm the coding processes for multiple (at least 3 or 4) ways to determine each parameter.

Once you have a list of ideas, select the two most promising approaches for each parameter. Think about those approaches in both a conceptual way (describe them in words, flowcharts, and/or sketches) and a practical way (how you will implement it in MATLAB). Write clear steps for identifying each parameter that can be translated into MATLAB code – do not write code. Consider providing a flowchart to clarify your approaches. Use Word’s draw tools or hand sketches to clarify your approach as necessary.

In Tables 2, 3, and 4 below, describe your two approaches for identifying each parameter: v_{0_i}, V_{max}, and K_m. For each,

- briefly and clearly explain your approach using steps. You can also include other items to clarify such as flowchart, pseudocode, or neat sketches;
- provide an evidence-based justification for each approach, keeping in mind that the parameters must be identified in a fully automated way, without any user intervention. Your justifications must make explicit reference to MATLAB functions and coding techniques needed to translate your steps to operational code;
- mark the expected level of difficulty (low, medium, or high) for implementing this approach in MATLAB; then justify in 1-2 sentences your selected level of difficulty, using evidence-based rationales if possible.

In Table 5, cite any references that helped you formulate your approaches. You will need at least one citation for each of Table 2, Table 3, and Table 4; however, you will likely need more than one citation per table. Use APA format. Remember to use in-text citations as well.

Table 2. Two coding approaches to determine v_{0_i}

Approach #1 for Parameter v_{0_i}
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Take the data and graph over the x values t = 0 to t = 100 2. Find the line of best fit of the graph 3. Create a function to find slope of t 4. Calculate slope for the first part of the graph. (Add more steps as needed)
What is your evidence-based justification for your approach? If the graph is over a very low amount of x, relative to the entire data set then taking the slope of that graph would be very close to taking the instantaneous velocity and could give us an accurate slope of the line.
Expected difficulty for coding in MATLAB: ___ low _x__ medium ___ high Why do you think it will be at the level of difficulty you indicated? We know how to do most of the needed code, however it may be tough to make the graph
Approach #2 for Parameter v_{0_i} (Must be different from Approach #1)
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Import the data from the .csv file 2. Find the length of the test using length function 3. Find 5% of the length 4. Find the slope between the point t = 0 and t = 5% of length to get v0

(add more steps as needed)
What is your evidence-based justification for your approach? This will give us the slope of the line created by the points at time = 0 and time = 5% of length. This would give us the slope of the beginning 5% of the curve.
Expected difficulty for coding in MATLAB: _x__ low ___ medium ___ high Why do you think it will be at the level of difficulty you indicated? We will only need to code one equation that does not include too many different variables or operations. It is just coding the slope equation with our data and then displaying it so it will not be too much of a struggle.

Table 3. Two coding approaches for determining V_{max}

Approach #1 for Parameter V_{max}
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Create a Michaelis-Menten graph using the $\boldsymbol{v_{0_i}}$ we found above. 2. Use the max function on the y data of the Michaelis-Menten graph.
What is your evidence-based justification for your approach? We are trying to find the maximum velocity and if we have a dataset of all of the velocities at every data point, we can use the max function to find the max velocity.
Expected difficulty for coding in MATLAB: __x_ low ___ medium ___ high Why do you think it will be at the level of difficulty you indicated? It will only be a few lines of code with only one function.
Approach #2 for Parameter V_{max} (Must be different from Approach #1)
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Graph the data 2. Find the line of best fit of the graph 3. Find the max of the equation of the line of best fit over the x values (add more steps as needed)

Expected difficulty for coding in MATLAB: ___ low ___ medium __x_ high Why do you think it will be at the level of difficulty you indicated? It is a multistep process with elements we haven't quite experienced in matlab before.

Table 4. Two coding approaches for determining K_m

Approach #1 for Parameter K_m
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Divide Vmax by 2 2. loop through all of the indices to find where the desired value is in between 2 of the y values 3. Make the higher of the corresponding x values the Km value (add more steps as needed)
What is your evidence-based justification for your approach? Once we have our desired value, we can make a loop to check where in the data our value it. Once we know what two values the Km lies between, we can assign the Km to the higher value
Expected difficulty for coding in MATLAB: ___ low __x_ medium ___ high Why do you think it will be at the level of difficulty you indicated? This approach involves loops and is unfamiliar to us.
Approach #2 for Parameter K_m (Must be different from Approach #1)
Approach (words, flowcharts, sketches) Steps, in English, not code: 1. Divide the Vmax by 2 2. Create an error range around our desired value 3. Use the find function to find all data points within the range of data 4. Look at the data and see what data point has the closest value to vmax 5. Use the corresponding x value as the Km value (add more steps as needed)
What is your evidence-based justification for your approach? When we find the desired value, we do not know if any of our data points will have that exact y value. So instead of searching for that specific value, we should find all data points within a small range close to that value and then look at them to see which is the closest to the vmax/2. Once we find that point, we can use that x value for the km value.
Expected difficulty for coding in MATLAB: ___ low ___ medium _x__ high

Why do you think it will be at the level of difficulty you indicated?
This is a complicated process that may take multiple attempts to figure out how to make it work just right.

Table 5. References Used in Evidence-Based Rationales (Part 2, Section A)

Purdue FYE – Ideas to Innovation. (2021, March 24). ENGR 132 – Project – Milestone 1a [Video]. YouTube. https://www.youtube.com/watch?v=qG62kX-8hrg

McFall, G (2022) Class 11B: Milestone 1 Overview [PowerPoint slides].

College of Engineering, Purdue University

Section B: Evaluate Processes Using Experimental Data

For this project, NaturalCatalysts measured the enzyme performance of 5 next-generation enzymes. For each enzyme, they measured the change in product concentration over time in kinetic-enzyme tests for 10 different substrate concentrations, $[S_i]$. Also, each test was repeated. So, you have a total of 100 tests, 20 for each enzyme. This real experimental data contains measurement noise, an example of which is shown in Figure 1.

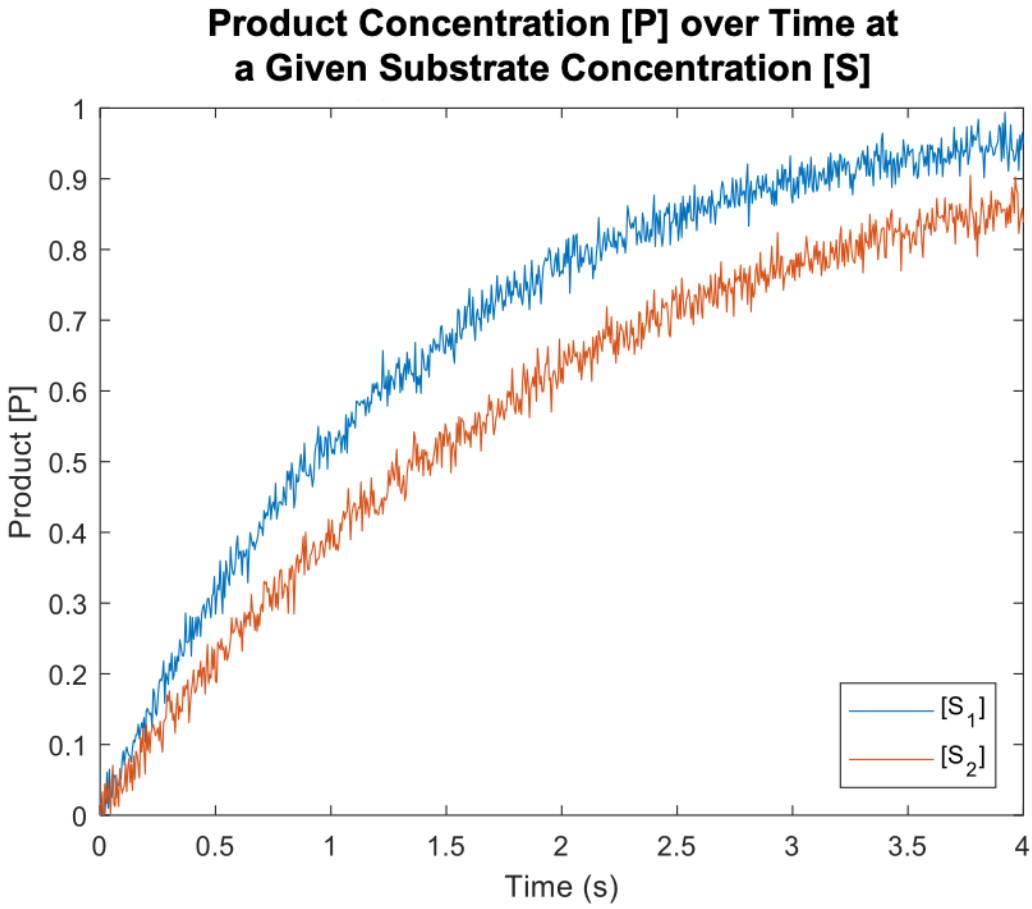

Figure 1. *Two examples of conversion of substrate to product over time during an enzyme reaction with noise present in the data.*

Experimental noise can come from many different sources, and it always makes parameter identification more challenging. In this part of the assignment, you must re-evaluate your approaches from Part 2, Section A in light of the measurement noise.

For each of the approaches you described in Table 2:

- explain in a few sentences how the approach might fail in the presence of noise. As appropriate, include sketches to clarify your explanation.
- modify your approach to account for the presence of noise in the reaction data. Use flowcharts and sketches to clarify your steps as necessary.
- re-assess the difficulty of coding this approach in MATLAB for analysis of noisy data and then justify in a few sentences your selected level of difficulty using evidence-based rationales (remember, your parameter identification must be fully automated). Again, your justifications must make explicit reference to MATLAB functions and coding techniques needed to translate your steps to operational code.

In Table 8, cite any references that helped you formulate your approaches. You will need at least one citation for Table 6. Table 7 will likely also need citations. However, you will likely need more citations in

this section than the minimum listed here. You are allowed to re-use the citations from Part 2a in this section, provided they make sense. Use APA format. Remember to use in-text citations as well.

Table 6. Evaluate the effect of noisy data on coding approaches

Evaluate Approach #1 for Parameter v_{0_i} (Refer to Approach #1 from Table 2)
How could this approach fail in the presence of noise? Noise can tweak the data to move the value of the parameter up or down depending on how far the outliers stray from the general trend. However, the overall method will not fail with the presence of noise. We will be taking the best fit line of the first part of the graph which will account for outliers and then calculating the slope of the best fit line.
Modify this approach to account for noise. Steps, in English, not code: 1. N/A 2. (add more steps as needed)
What is your evidence-based justification for your approach? Because we are already using the best fit line of the data for the calculation, it should not matter if there is noise in the graph. The noise will affect the values, but not the method of obtaining the v0i
Expected difficulty for coding in MATLAB: _x__ low ___ medium ___ high Why do you think it will be at the level of difficulty you indicated? We will not have to modify this approach in the presence of noise because it already accounts for noisy data. Because there will be no modifications, it should not be too difficult.
Evaluate Approach #2 for Parameter v_{0_i} (Refer to Approach #2 from Table 2)
How could this approach fail in the presence of noise? This code could fail in the presence of noise if the data point we use at 5% of the length happens to be a major outlier and the calculated slope does not represent the actual v0i
Modify this approach to account for noise. Steps, in English, not code: 1. Use the matlab function isoutlier to determine if the point at 5% of the length is an outlier 2. Perform the approach on the non-oultliers after filtering the data

(add more steps as needed)
What is your evidence-based justification for your approach? isoutlier exists in the MATLAB documentation, so there is no doubt the approach can help filter the outliers
Expected difficulty for coding in MATLAB: ___ low __x_ medium ___ high Why do you think it will be at the level of difficulty you indicated? Filtering through a large dataset may take some work, but the built-in function helps the difficulty of the approach.

Consider how the noise and the updated approaches to v_{0_i} could affect your approach to finding V_{max} (from Table 3) or K_m (from Table 4). Explain any modifications needed, or explain why no modifications are needed.

Table 7. Modifications and justifications to V_{max} or K_m approaches

Updates to V_{max} approaches
No modifications would be needed, if our calculated v0's are accurate then our calculations for Vmax should not be affected. Vmax is measured from the Michaelis-Menten graph. The Michaelis-Menten graph is a graph of v0 vs concentration, so if v0 is accurate then our Vmax will be accurate.
Updates to K_m approaches
No modifications would be needed, if our calculated v0's are accurate then our calculations for Km should not be affected. Km is measured from the Michaelis-Menten graph. Specifically it is the concentration when the velocity is half of Vmax, so if the v0's are calculated correctly and not affected by the noise then our Km should not be affected either.

In Table 8, cite any references that helped you formulate your approaches. Use APA format. Remember to use in-text citations as well.

Table 8. References Used in Evidence-Based Rationales (Part 2, Section B)

Find outliers in data - MATLAB isoutlier - MathWorks. MATLAB. (n.d.). Retrieved March 29, 2022, from https://www.mathworks.com/help/matlab/ref/isoutlier.html
Purdue FYE – Ideas to Innovation. (2021, March 24). ENGR 132 – Project – Milestone 1a [Video]. YouTube. https://www.youtube.com/watch?v=qG62kX-8hrg

McFall, G (2022) Class 11B: Milestone 1 Overview [PowerPoint slides]. College of Engineering, Purdue University

How to Submit

1. Rename this answer sheet to be **M1_AnswerSheet_*SSS*_*TT*.docx** 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 the answer sheet as one PDF named M1_***SSS*_*TT*.pdf**.
3. Select one person to submit the PDF for the team. That person should
    a. Log into Gradescope and submit **M1_AnswerSheet_*SSS*_*TT*.pdf** to the **M1** assignment.
    b. Select all team members for the group assignment.
    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 and that all parts of the answer sheet were properly tagged.
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

Idea Fluency (IF)

Generate ideas fluently. Take risks when necessary.

IF01. Generate a wide range of solutions including ideas not readily obvious or combinations of ideas in new ways.

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.

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.