MTE54 Fall 2024
Assignment 2

These guidelines are designed to help students present their work clearly and professionally, enabling graders to provide accurate and timely feedback. With over a hundred assignments to review, a consistent format and organized presentation allow graders to focus on evaluating each solution rather than adjusting to different styles. By following these standards, students support a smoother grading process that benefits everyone, making it easier to receive prompt and meaningful feedback.

Code Submission Guidelines

1. Ensure code is properly indented, with comments explaining the purpose and logic of each section.
2. Follow the naming convention specified in each template, noted at the top of the code.
3. Every programming-related files including the dataset must be compressed as a single zip-file.

Typed Assignment Guidelines (Preferred)

1. Clearly label each section and problem with appropriate headings and subheadings.
2. Use LaTeX or a built-in equation editor for all mathematical expressions to maintain clarity.
3. Submit typed assignments in PDF format to prevent formatting issues (Word documents are not accepted).
4. This pdf file must not be include in the programming zip file.
5. Embed any required diagrams or plots, with clear captions or labels.
6. Choose legible fonts in terms of size, style, and color.
7. Start a new page for each problem to enhance clarity and organization.

Hand-Written Assignment Guidelines

1. If any part of the handwriting is deemed illegible, that section will not be graded.
2. Only black or blue ink pens are allowed; HB pencils are not permitted.
3. Assignments must be neatly organized, with clear headings, subheadings, and numbered sections.
4. Text should flow from left to right and top to bottom on each page.
5. Use standard 8.5”x11” lined or grid paper only.
6. Maintain at least a 1-inch margin on all sides of each page.
7. Begin a new page for each problem.
8. Number pages visibly at the bottom to avoid misordering.
9. Include your name on the first page of the assignment.
10. All page must be converted to a single pdf file.

Note

Failure to comply will result in a mark deduction for each violation.

Please use computer-typed equations and text for the entire assignment, avoid hand-writing when possible.

1 RANSAC with Ellipse I (18)

In this exercise, you are tasked to detect an ellipse based on given dataset using RANSAC.

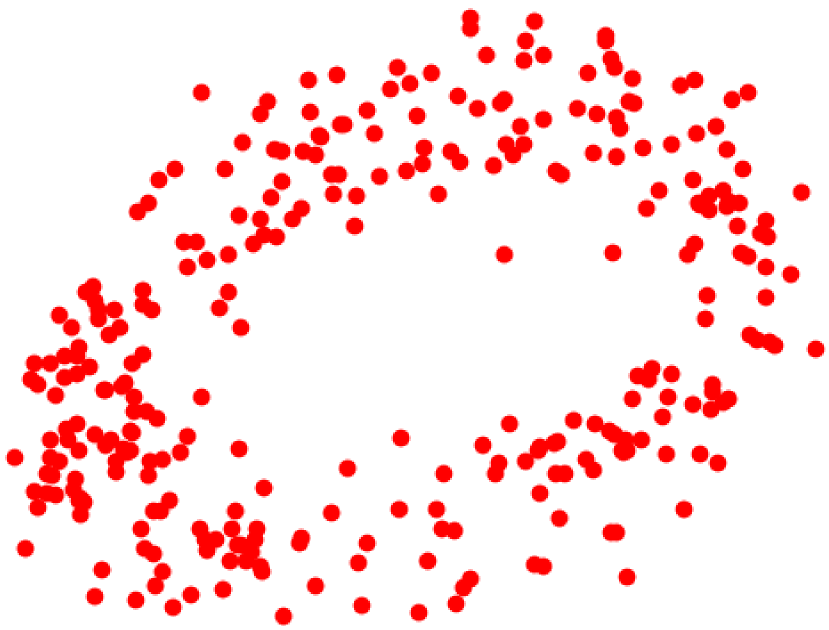

Figure 1: A Plot of Randomly Selected Dataset

The goal is to estimate the parameters $\mathbf{Q}$ and $\mathbf{b}$ by fitting the sensor data to an ellipse. An ellipse can be described by the quadratic equation:

$$(\mathbf{p}-\mathbf{b})^\top \mathbf{Q}(\mathbf{p}-\mathbf{b}) = 1$$

or

$$Ax^2 + 2Bxy + Cy^2 - 2Dx - 2Ey + 1 = 0$$

where:

- $\mathbf{Q} = \mathbf{Q}^\top \in \mathbb{R}^{2\times 2}$ is a unique symmetric positive definite matrix of coefficients. [If $\mathbf{Q}$ is not positive definite, you may have a hyperbola instead.]

Note that the coefficients (A,B,C,D,and E) are also unique. The conversion from the polynomial coeffcients to the quadratic form can be expressed as follows:

$$\mathbf{Q} = \alpha \begin{bmatrix} A & B \\ B & C \end{bmatrix}$$

$$\mathbf{Qb} = \alpha \begin{bmatrix} D \\ E \end{bmatrix}$$

$$\alpha = \frac{1}{\begin{bmatrix} D & E \end{bmatrix} \begin{bmatrix} A & B \\ B & C \end{bmatrix}^{-1} \begin{bmatrix} D \\ E \end{bmatrix} - 1}$$

a.) Determine the minimum number of points to uniquely determined ellipse's parameters as well as derive a method to construct an ellipse with the parameters (A,B,C,D, and E) based on those points. $\mathbf{p}_i = (x_i, y_i)$ denote the i^{th} data point. You must explain your rationale to receive credits. [10 pts]

Answer: The general form of an ellipse is:

$$Ax^2 + 2Bxy + Cy^2 - 2Dx - 2Ey + 1 = 0$$

From this equation, it is evident that with five unknowns, five points are needed to uniquely determine an ellipse's parameters. Each point is written as:

$$p_i = (x_i, y_i)$$

Substituting $x = x_i$ and $y = y_i$ for $i = 1, 2, 3, 4, 5$ into the ellipse equation. This results in a system of five linear equations:

$$\begin{cases} Ax_1^2 + Bx_1y_1 + Cy_1^2 + Dx_1 + Ey_1 = 1 \\ Ax_2^2 + Bx_2y_2 + Cy_2^2 + Dx_2 + Ey_2 = 1 \\ Ax_3^2 + Bx_3y_3 + Cy_3^2 + Dx_3 + Ey_3 = 1 \\ Ax_4^2 + Bx_4y_4 + Cy_4^2 + Dx_4 + Ey_4 = 1 \\ Ax_5^2 + Bx_5y_5 + Cy_5^2 + Dx_5 + Ey_5 = 1 \end{cases}$$

Rewriting the above system of linear equations:

$$\begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

Solving:

$$\begin{bmatrix} A \\ B \\ C \\ D \\ E \end{bmatrix} = \begin{bmatrix} x_1^2 & x_1y_1 & y_1^2 & x_1 & y_1 \\ x_2^2 & x_2y_2 & y_2^2 & x_2 & y_2 \\ x_3^2 & x_3y_3 & y_3^2 & x_3 & y_3 \\ x_4^2 & x_4y_4 & y_4^2 & x_4 & y_4 \\ x_5^2 & x_5y_5 & y_5^2 & x_5 & y_5 \end{bmatrix}^{-1} \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

b.) Fit an ellipse with the following datasets and compute for $\mathbf{Q}$, $\mathbf{b}$, A, B, C, D and E.

$$(2.92, -6.01), \quad (3.40, -7.20), \quad (4.99, -7.84), \quad (5.48, -7.04), \quad (4.20, -5.91)$$

You don't have code for this part, but you still have to do it with Python in the next part. [8 pts.]

2 RANSAC with Ellipse II (30)

In this exercise, you are tasked to implement RANSAC algorithm for determining the shape of an ellipse in Python. The implementation must match your explanation in the report in order to receive full credits.

To test whether point $\mathbf{p}_i$ is an inlier, let the following inequality be the condition for inliers.

$$\left\| \sqrt{(\mathbf{p}_i - \mathbf{b})^\top \mathbf{Q} (\mathbf{p}_i - \mathbf{b})} - 1 \right\| \leq \varepsilon$$

where ε is the tunable threshold.

a.) Based on the given template in Python, implement *fit_ellipse_subset* that compute $\mathbf{Q}$ and $\mathbf{b}$ from a given set of points. (The number points is determined from 1.a). [8 pts]

b.) Based on the given template in Python, implement *ransac_ellipse* that determined the best $\mathbf{Q}$, $\mathbf{b}$, and the corresponding inliers based on the given dataset, number of iterations, and threshold. You must explain your implementation step-by-step [12 pts] with the rationale in your report.

c.) Based on the given template in Python, use different number of dataset, number of iterations, and threshold value. Comment on the effect of these parameters. [10 pts]

3 Bayesian Filter for Tracking Moo-Deng's Behavior I (32)

Figure 2: An abstract version of Moo-Deng's enclosure

Khao Kheow Open Zoo in Thailand has gained popularity thanks to its star resident, Moo-Deng, a playful yet feisty Pygmy hippo. To better understand Moo-Deng's daily routine, the zoo has hired a wildlife researcher to track her movements between three key zones: Feeding Area ($A = 1$), Resting Area ($A = 2$),Playground Area ($A = 3$). Let $\mathbf{x}$ be a vector of probability of Moo-Deng being in the feeding area, the resting area, and the playground area respectively, which can be formally expressed as follows:

$$\mathbf{x} = \begin{bmatrix} p(A = 1) \\ p(A = 2) \\ p(A = 3) \end{bmatrix}$$

It can be assumed that Moo-Deng only stays in one of these 3 places. Therefore:

$$\sum_{i=1}^{3} p(A = i) = \mathbf{1}^\top \mathbf{x} = 1$$

After careful observation, the researcher has identified Moo-Deng's behavior patterns:

- When Moo-Deng is in the Feeding Area ($A = 1$), she has a 60% chance of staying there, and a 40% chance of moving to the Resting Area ($A = 2$).
- If Moo-Deng is in the Resting Area ($A = 2$), she has a 40% chance of staying there, a 20% chance of returning to the Feeding Area ($A = 1$), and a 40% chance of moving to the Playground Area ($A = 3$).
- If Moo-Deng is in the Playground Area ($A = 3$), she has a 70% chance of staying there, a 20% chance of moving to the Feeding Area ($A = 1$), and a 10% chance of moving to the Resting Area ($A = 2$).

The behavior of Moo-Deng can be modelled as a transition matrix $\mathbf{A} \in \mathbb{R}^{3\times 3}$ such that the following is true:

$$\mathbf{x}_{k+1} = \mathbf{A}\mathbf{x}_k$$

where $\mathbf{x}_k$ denotes the probability vector at the k^{th} observation.

Furthermore, the researcher uses a digital sensor system to track Moo-Deng's location, but the sensor is not perfectly accurate:

- When Moo-Deng is in the Feeding Area, the sensor reports Feeding Area with an 80% accuracy, but it incorrectly reports Resting Area 10% of the time and Playground Area 10% of the time.
- When Moo-Deng is in the Resting Area, the sensor reports Resting Area 70% of the time, but it incorrectly reports Feeding Area 20% of the time and Playground Area 10% of the time.
- When Moo-Deng is in the Playground Area, the sensor reports Playground Area with an 85% accuracy, but it incorrectly reports Resting Area 10% of the time and Feeding Area 5% of the time.

At the k^{th} observation, the reading from the sensor system can be encoded to a vector of digital values $\mathbf{y}_k$ so that:

$$\mathbf{y}_k = \begin{cases} \begin{bmatrix}1 & 0 & 0\end{bmatrix}^\top & \text{if} \quad \text{the measurement displays } \texttt{Feeding Area} \\ \begin{bmatrix}0 & 1 & 0\end{bmatrix}^\top & \text{if} \quad \text{the measurement displays } \texttt{Resting Area} \\ \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\top & \text{if} \quad \text{the measurement displays } \texttt{Playground Area} \end{cases}$$

Then, the probaiblity of sensor displaying each area can be modelled as follows:

$$\begin{bmatrix} p(\mathbf{y}_k = \begin{bmatrix}1 & 0 & 0\end{bmatrix}^\top) \\ p(\mathbf{y}_k = \begin{bmatrix}0 & 1 & 0\end{bmatrix}^\top) \\ p(\mathbf{y}_k = \begin{bmatrix}0 & 0 & 1\end{bmatrix}^\top) \end{bmatrix} = \mathbf{C}\mathbf{x}_k$$

where $\mathbf{C} \in \mathbb{R}^{3\times 3}$ is yet to be determined.

a) Assume that Moo-Deng starts in the Feeding Area ($A_0 = 1$). What is the probability that her subsequent movements will follow the sequence (2,3,3,1) in this exact order? Explain your calculation. [4 pts]

b) Determine the numerical value of the transition matrix $\mathbf{A}$. You must explain your rationale. [5 pts]

c) Analytically compute the stationary probabilities of Moo-Deng's location $\mathbf{x}^*$. These are the probabilities that, after many transitions, Moo-Deng will be found in each of the zones regardless of the initial location. You must show your work. [Hint: ?? = $\mathbf{A}\mathbf{x}^*$] [8 pts]

d) Determine the numerical value of the matrix $\mathbf{C}$. You must explain your rationale. [5 pts]

e) Derive a state estimator based on a Bayesian Filter that estimates the states and probability of Moo-Deng's based on prior belief $\mathbf{x}_{k-1}$ and the current sensor measurement $\mathbf{y}_k$. You may leave your answer in terms of $\mathbf{A}$ and $\mathbf{C}$. You may also substitute their numeric values. [10 pts]

4 Bayesian Filter for Tracking Moo-Deng's Behavior II (20)

In this exercise, you are tasked to implement a state estimator for Moo-Deng's whereabouts along with a simple behavior simulator in Python . The implementation must match your explanation in the report in order to receive full credits.

a) Based on the given code in Python, complete the function `moodeng_behavior_update` that randomly returns the state of Moo-Deng based on the given current state and the transition matrix $\mathbf{A}$. You must explain your implementation in this report. Comment in the report on how can you use statistics to verify your results ? [5 pts.]

b) Based on the given code in Python, complete the function `sensor_measurement` that return the noisy measurement $\mathbf{y}_k$ based on the matrix $\mathbf{C}$ and given Moo-Deng's state. Note that the results should be either (1,0,0), (0,1,0), or (0,0,1). [5 pts.]

c) Based on the designed Bayesian filter and the given code in Python, implement a state estimator and the rest of the simulation in called `sim_moodeng`. The estimator should :

- estimate a sequence state of Moo-Deng's location $\hat{A}_k$
- return a sequence belief (probability vector) of Moo-Deng's location $\mathbf{x}_k$

given an initial belief $\mathbf{x}_0$ and a sequence of sensor measurement $\mathbf{y}_k$ [6 pts.].

Submission Guideline

Include details of all analytical computations and assumptions made (if any) and required plots in your assignment report. There is no page limit. Comment your code appropriately. Submit a single PDF for the report and a single zip file with all the codes on Dropbox by Nov 8th, 11:59PM. Late penalty applies as per syllabus if not properly justified. Overlaps with other assignments and/or course duties are not accepted as justifications.