NA 568 Mobile Robotics: Methods & Algorithms Winter 2020 – PS1

Maani Ghaffari
University of Michigan

January 9, 2020

This problem set counts 10% of your course grade. You are encouraged to talk at the conceptual level with other students, but you must complete all work individually and may not share any non-trivial code or solution steps. See the syllabus for the full collaboration policy.

Submission Instructions

Your assignment must be received by 11:55 pm on Sunday, January 26. You are to upload your assignment directly to the Gradescope website as two attachments:

1. A `.tar.gz` or `.zip` file *containing a directory* named after your uniqname with the structure shown below.

   ```
   alincoln_ps1.tgz:
   alincoln_ps1/
   alincoln_ps1/task2c.m
   alincoln_ps1/task3.m
   alincoln_ps1/task4c.m
   ```

2. A PDF with the written portion of your write-up. Scanned versions of hand-written documents, converted to PDFs, are perfectly acceptable. No other formats (e.g., .doc) are acceptable. Your PDF file should adhere to the following naming convention: `alincoln_ps1.pdf`.

Homework received after 11:55 pm is considered late and will be penalized as per the course policy. The ultimate timestamp authority is the one assigned to your upload by Gradescope. No exceptions to this policy will be made.

Important: For all (x, y) plots you will want to use the `axis equal` command to set the aspect ratio so that equal tick mark increments on the x-,y- and z-axis are equal in size. This makes `SPHERE(25)` look like a sphere, instead of an ellipsoid.

Task 1: Probability Basics (25 points)

A. (5 pts) State the definition of $\mathbb{E}[X + Y + Z]$ in terms of the underlying probability distribution, where the random variables are continuous. Then prove that $\mathbb{E}[X + Y + Z] = \mathbb{E}[X] + \mathbb{E}[Y] + \mathbb{E}[Z]$, using the definition of expectation. Your proof should be succinctly and clearly written. **Do NOT assume that the random variables are independent.**

B. (5 pts) For each of the covariance matrices below, indicate whether it is Valid or Invalid. If it is invalid, give a reason.

(a) $\begin{bmatrix} 8 & -5 \\ -5 & 10 \end{bmatrix}$

(b) $\begin{bmatrix} 3 & 7 \\ 7 & -5 \end{bmatrix}$

(c) $\begin{bmatrix} 8 & 5 \\ 3 & 3 \end{bmatrix}$

(d) $\begin{bmatrix} 3 & 7 \\ 7 & 4 \end{bmatrix}$

(e) $\begin{bmatrix} 5 & 4 \\ 4 & 8 \end{bmatrix}$

C. (8 pts) Rick has collected four points (in 2D space); let's call this set 'r'. He has computed their mean and (biased) sample covariance to be:

$$\mu_r = \begin{bmatrix} 2 \\ 3 \end{bmatrix} \quad \Sigma_{rr} = \begin{bmatrix} 5 & 3 \\ 3 & 7 \end{bmatrix}$$

And Morty has collected six points (from the same 2D space); let's call it set 'm', finding their mean and (biased) sample covariance to be:

$$\mu_m = \begin{bmatrix} -2 \\ 2 \end{bmatrix} \quad \Sigma_{mm} = \begin{bmatrix} 8 & 4 \\ 4 & 3 \end{bmatrix}$$

(a) If Rick computed the sum (over all r points) of $xx^\top$, what value would he have computed? (Show work and provide a numerical answer.)

(b) If Morty computed the sum (over all m points) of $xx^\top$, what value would he have computed? (Show work and provide a numerical answer.)

We now wish to compute the mean and sample covariance of all ten points (which we will denote $m + r$).

(c) What is the mean, μ_{m+r}? (Show work and provide a numerical answer.)

(d) What is the (biased) sample covariance, Σ_{m+r}? (Show work and provide a numerical answer.)

D. (7 pts) Answer the following. Any proofs should be succinctly and clearly.

(a) Prove that if A and B are independent events, so are A^C and B^C. A^C is the complement of A such that A $\cap A^C = \varnothing$.

(b) Prove that the affine transformation of a Gaussian random vector is Gaussian.

Task 2: Bayes Filter (25 points)

A. (7 pts) Suppose that 1% of the general population has cancer and that a particular test for cancer has a 20% false positive rate and a 10% false negative rate. Your test result is positive. What is the probability that you indeed have cancer?

Suppose now that the National Institute of Health has $1,000,000 of research funding to invest in improving this test. In this particular instance which would have more bang for their research buck and why—reducing the false positive rate or reducing the false negative rate?

B. (8 pts) A robot is equipped with a manipulator to paint an object. Furthermore, the robot has a sensor to detect whether the object is colored or blank. Neither the manipulation unit nor the sensor are perfect.

From previous experience you know that the robot succeeds in painting a blank object with a probability of

$$p(x_{t+1} = \text{colored} \mid x_t = \text{blank}, u_{t+1} = \text{paint}) = 0.9,$$

where x_{t+1} is the state of the object after executing a painting action, u_{t+1} is the control command, and x_t is the state of the object before performing the action.

The probability that the sensor indicates that the object is colored although it is blank is given by $p(z = \text{colored} \mid x = \text{blank}) = 0.2$, and the probability that the sensor correctly detects a colored object is given by $p(z = \text{colored} \mid x = \text{colored}) = 0.7$.

Unfortunately, you have no knowledge about the current state of the object. However, after the robot performed a painting action the sensor of the robot indicates that the object is colored.

Compute the probability that the object is blank after the robot has performed an action to paint it. Use an appropriate prior distribution and justify your choice.

C. (10 pts) Consider a robot that resides in a circular world consisting of ten different places that are numbered counterclockwise. The robot is unable to sense the number of its present place directly. However, places 0, 3, and 6 contain a distinct landmark, whereas all other places do not. All three of these landmarks look alike. The likelihood that the robot observes the landmark given it is in one of these places is 0.8. For all other places, the likelihood of observing the landmark is 0.4.

For each place on the circle we wish compute the probability that the robot is in that place given that the following sequence of actions is carried out deterministically and the following sequence of observations is obtained: The robot detects a landmark, moves 3 grid cells counterclockwise and detects a landmark, and then moves 4 grid cells counterclockwise and finally perceives no landmark.

Implement the circular world described above using a discrete Bayes filter in Matlab to numerically arrive at the desired belief.

Task 3: First-Order Covariance Propagation (25 points)

In most SLAM approaches, the noise characteristics of individual observations are linearized, resulting in a Gaussian covariance matrix. In this task, we explore the effects of the error introduced via linearization.

Consider a robot sensor that observes range and bearing to nearby landmarks. In this case, the range error is relatively small, but the bearing error is large. We are interested in determining the (x, y) position of the beacon based on observations obtained by the robot at the origin $(0, 0)$. For this problem, the robot does not move.

Suppose you obtain a (range, bearing) observation with mean $(10.0\,\text{m}, 0\,\text{rad})$ whose range standard deviation is $0.5\,\text{m}$, and whose bearing standard deviation is $0.25\,\text{rad}$. The range and bearing measurements are Gaussian and independent.

A. (4 pts) Generate a point cloud representing 10,000 samples from the distribution over the position of the beacon as measured in

i) the sensor frame, i.e. (r, θ) space and

ii) the Cartesian (x, y) coordinate frame.

In other words, generate observations of (range, bearing) and project these points into (x, y). (Hint: use the `randn()` function in Matlab, and recall that you can sample from a univariate Gaussian with mean `mu` and standard deviation `sigma` with `mu + sigma*randn`.)

B. (4 pts) What is the (linearized) covariance of the beacon position in (x, y) coordinates? In other words, write the covariance of an observation in (x, y) coordinates in terms of the covariance of the observation in (range, bearing) coordinates. The transformation is non-linear, so you will need to compute a first-order approximation (Taylor expansion) of the transformation function. Make the appropriate Jacobians easy to read in your source code, using comments if necessary.

C. (4 pts) Draw in red the 1-sigma, 2-sigma, and 3-sigma contours of the analytical (linearized) covariance ellipses, super-imposed over the point clouds generated in parts A.i and A.ii. Now overlay in blue the actual covariance ellipses computed using sample-based expressions for the first and second moments. Do they agree? Why or why not?

You may use the function `draw_ellipse()` provided in the `ps1-code.zip`. For its syntax, do `help draw_ellipse` at the command prompt.

D. (4 pts) From a purely theoretical perspective, assuming that the underlying process is truly Gaussian, we expect 39.35% of all samples to lie within the 1-sigma contour, 86.47% of samples to lie within the 2-sigma contour, and 98.89% to lie within the 3-sigma contour. (These frequencies were computed using the cumulative chi-square distribution for two degrees-of-freedom, i.e. χ^2_2. In Matlab, you can do this using the function `chi2cdf()` evaluated at `chi2cdf(1,2)`, `chi2cdf(4,2)`, and `chi2cdf(9,2)`, respectively.)

Modify your software to count the samples falling within each (analytical) ellipse for parts A.i and A.ii. The error of a particular sample x, measured in "units" of sigma, is known as the Mahalanobis distance, and can be computed as $\sqrt{(x-\mu)^\top \Sigma^{-1}(x-\mu)}$.

E. (4 pts) If the point samples were truly distributed as a Gaussian (clearly the (x, y) are not), your counts would match the theoretically predicted values. Try varying the noise parameters: under what conditions do the counts come close to matching the theoretically predicted values? What consequences to a state estimation algorithm could these sorts of errors have?

F. (5 pts) Suppose now that the (range, bearing) measurements are *not independent* but instead jointly correlated under the following three scenarios: a) $\rho_{r\theta} = 0.1$, b) $\rho_{r\theta} = 0.5$, and c) $\rho_{r\theta} = 0.9$. Repeat parts A and C. (Hint: use the `chol()` function in Matlab.)

Task 4 (25 points)

Assume that we want to estimate an unobserved population parameter θ on the basis of observations x. Let f be the sampling distribution of x so that $f(x|\theta)$ is the probability of x when the underlying population parameter is θ. The function $L(\theta) = f(x|\theta)$ when viewed as a function of the parameter θ is called the likelihood function or just the likelihood. For example, if x follows a Gaussian distribution, we will have $\theta = (\mu, \sigma)$ and

$$\mu, \sigma \mapsto f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)$$

is the likelihood.

Maximum Likelihood Estimator (MLE): The maximum likelihood method maximizes the likelihood function, leading to the MLE

$$\hat{\theta}_{MLE} = \arg\max_{\theta} L(\theta) = \arg\max_{\theta} f(x|\theta).$$

Maximum A Posteriori (MAP) Estimator: In the Bayesian framework, one can place a prior distribution over the parameter, i.e., $g(\theta)$. Then, the MAP estimator maximizes the posterior probability density function $f(\theta|x)$ as follows.

$$\hat{\theta}_{MAP} = \arg\max_{\theta} f(\theta|x) = \arg\max_{\theta} \frac{f(x|\theta)g(\theta)}{\int_{\Theta} f(x|\vartheta)g(\vartheta)d\vartheta} = \arg\max_{\theta} f(x|\theta)g(\theta),$$

were the last equality is true because the normalization constant in the Bayes' formula is independent of θ.

Remark 1. *Since* log *is a monotonic function, it is often the case that we use the logarithm of the likelihood or posterior for maximization (or negative of the logarithm for minimization).*

Now suppose we have a continuous random variable $\theta \sim \mathcal{N}(\mu, \sigma^2)$. We wish to infer its mean and variance as we obtain normally distributed measurements sequentially. For the case of a random mean, μ, and fixed variance, σ^2:

A. (5 pts) Derive a formula that provides a point-estimate of the posterior θ using a MAP estimator. Place a prior over the random variable as $\theta \sim \mathcal{N}(\mu_0, \sigma_0^2)$.

B. (5 pts + 5 Extra credits) Derive formulas to make a Bayesian inference so that we can infer both the mean and the variance. **Hint:** use Bayes' formula and substitute the Gaussian prior and likelihood in the formula.

C. (15 pts + 5 Extra credits) You are responsible for purchasing a sensor that can measure the range (distance) to an object. Sensor I ($100) and II ($500) are both used to measure the range to an object. Suppose the measurements are noisy values of the range, x, such that $z \sim \mathcal{N}(\mu_z, \sigma_z^2)$ with variances of 1 (I) and 0.64 (II). The measurements obtained from these sensors can be seen in Table I and II. Parameterize the prior of x with $\mu = 0$ and $\sigma^2 = 1000$. Using the derivations from part B, write a Matlab function that takes data as input and solves the inference recursively.

   C.1 Use the sensor data and the Matlab function to infer the mean and variance of the normally distributed random variable x conditioned only on z_1.

   C.2 Use the sensor data and the Matlab function to infer the mean and variance of the normally distributed random variable x conditioned only on z_2.

   C.3 Why is it that x is more precise [1] when conditioned on z_1 even though sensor II is more accurate? which sensor do you recommend to be purchased?

[1] https://en.wikipedia.org/wiki/Precision_(statistics)

N	Z_1
1	10.6715
2	8.7925
3	10.7172
4	11.6302
5	10.4889
6	11.0347
7	10.7269
8	9.6966
9	10.2939
10	9.2127

Table 1: Sensor I Data

N	Z_2
1	10.7107
2	9.0823
3	9.1449
4	9.3524
5	10.2602

Table 2: Sensor II Data