Midterm Exam CS226

Stanford CS226 Statistical Algorithms in Robotics, Winter 2009/10

Full Name: ______________________

Email: ______________________

SOLUTIONS

Welcome to the CS223B Midterm Exam!

- The exam is 6 pages long. Make sure your exam is not missing any sheets. The exam has a maximum score of 100 points. You have 60 minutes.
- The exam is open book, open notes, but closed cell phones, etc.
- Write your answers in the space provided. If you need extra space, use the back of the preceding sheet.
- Write clearly and be concise.
- All points will be manually counted before certification.

Question	Points
1 (20 max)	
2 (20 max)	
3 (20 max)	
4 (20 max)	
5 (20 max)	
total	

1 Occupancy Grid Maps and Binary Bayes Filters 20pts

You learned about the log-odds update rule of binary Bayes filters.

1. What happens if we force the log-odds to lie inside an interval $[-b; b]$ for some positive, small b? In what type of situation would such a bound look like a plausible choice?

   Answer: It bounds the weight of prior evidence. This makes sense if the environment changes over time. Then past evidence cannot arbitrarily override new evidence. However, this is only a crude approximation to the correct answer that would result under a probabilistic next state model.

2. In the log-odds calculation, we have to constantly offset for the prior. Suppose we omit this term. Under what condition will this be mathematically justified? And, conversely, what is the worst thing that can happen when we drop this term?

   Answer: This is mathematically justified if the prior is 0.5. Worst case: Can easily arrive at the wrong answer with arbitrary certainty. The close the prior is to 0 or 1, the more pronounced the effect

3. In occupancy grid maps, any cell outside the field of view of a sensor is never updated. While this makes intuitive sense, we want you to prove that this follows from the basic log-odds update rule.

   Answer: With $p(x|z) = p(x)$ the measurement and prior update terms cancel each other out in Eq. (4.20) in the book. Just plug in $p(x)$ instead of $p(x \mid z)$ into (4.20) on page 96, and the corresponding teams cancel out.

2 Particle filters 20pts

1. Consider a robot whose motion is deterministic (=no motion noise). Will a particle filter (if implemented exactly, that is, without assumed motion noise) approximate the correct posterior?

   Answer: Yes! It does. However, the particle filter will perform poorly in practice unless we use a huge number of particles. One that adds fake motion noise may perform better.

2. Say we are resampling over and over again, without any motion update. What will happen in the limit?

   Answer: This filter will converge to a single particle.

3. What happens to the particle filter if we use exactly one particle, and why?

   Answer: The key insight is: measurements will be ignored. This is because in the resampling step, this one particle is picked regardless of the measurement probability. To see, realize that its weight will always be normalized to 1. (and yes, the particle filter won't work well, but this is not what we are asking here)

3 Gaussians and EKF 20pts

1. In class, we learned that the product of two Gaussians is once again Gaussian. Derive the variance ϕ^2 of the product of two 1-D Gaussians (don't derive the new mean).

$$\frac{1}{\sqrt{2\pi\,\sigma^2}}\exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2}\right\} \qquad \frac{1}{\sqrt{2\pi\,\rho^2}}\exp\left\{-\frac{1}{2}\frac{(x-\nu)^2}{\rho^2}\right\}$$

Answer: We did exactly this in class! The logarithm of this product is

$$\log p(x) = \text{const.} - \frac{1}{2}\left(\frac{(x-\mu)^2}{\sigma^2} + \frac{(x-\nu)^2}{\rho^2}\right)$$

This is obviously quadratic in x, as we discussed in class.
We get the new variance by taking the second derivative

$$\frac{\partial^2 \log p(x)}{\partial x^2} = \frac{1}{\sigma^2} + \frac{1}{\rho^2}$$

The second derivative of the resulting Gaussian is:

$$\frac{\partial^2 \log p(x)}{\partial x^2} = \frac{1}{\phi^2}$$

From that we get for the covariance

$$\phi^2 = (\sigma^{-2} + \rho^{-2})^{-1}$$

2. Prove that the resulting variance ϕ^2 you just derived is smaller than σ^2 and smaller than ρ^2 (feel free to ignore the degenerate case of infinite variance).

Answer: There are many different proofs, here is one. Develop

$$\begin{aligned}\phi^2 &= (\sigma^{-2} + \rho^{-2})^{-1}\\ &= \sigma^2(\sigma^2\sigma^{-2} + \sigma^2\rho^{-2})^{-1}\\ &= \sigma^2\left(1 + \frac{\sigma^2}{\rho^2}\right)^{-1}\\ &< \sigma^2\,(1)^{-1} = \sigma^2\end{aligned}$$

By symmetry, the same applies to ν^2.

3. In problems like localization, EKFs work poorly if the initial state is unknown (global localization), whereas they often work well if the initial state is known (tracking). Explain how this relates to the "E" in "EKF."

Answer: This is because of the linearization of the measurement model: the more spread the prior, the more inaccurate the approximation using a Taylor expansion. In the extreme case of global localization, the linearities simply won't allow for a meaningful prior. Put differently, the more spread the prior, the more multi-modal the posterior.

4 FastSLAM 20pts

The FastSLAM algorithm as discussed in class uses Gaussians to represent the 2-D landmark estimates (for notation: N denotes the number of path particles, and M the number of landmarks). Say we wish to replace those Gaussians by particle filters.

1. What would be the number of state variables in each of those new particle filters?

   Answer: 2, since each particle represents the location of exactly one landmark in 2-D.

2. How many such new particle filters would there be?

   Answer: $M \times N$, that is, each particle filter carries its own landmark estimates, and there will be M such landmarks.

3. When is this a good idea?

   Answer: This is a good idea when the measurement model is highly non-linear or degenerate, so that Gaussians won't be a good representation.

5 GraphSLAM 20pts

Say we are using GraphSLAM to build a 2-D map with K robots. For simplicity, we will assume that all measurement and motion equations are linear (so don't worry about linearization). Also, assume that all initial poses are known. Let us consider the case in which all robots communicate and maintain a single, joint belief.

1. How will the joint information matrix be structured? Draw a diagram of this matrix, and indicate which parts of this matrix will be sparse.

   Answer:

   $$\begin{pmatrix} I & \mathbf{sparse} \\ \mathbf{sparse}^{\mathbf{T}} & I \end{pmatrix} \tag{1}$$

   Here I is a diagonal matrix, and **sparse** *denotes a sparse matrix. Notice that there won't be any non-zero between robot elements in this matrix, and there won't be any zeros for the landmark-to-landmark elements either.*

2. Will it be symmetric? Will it be positive semi-definite?

   Answer: Yes, for both questions. This is always true about covariance matrices.

3. Say robots can sense each others relative distance; and again everything is linear. Does this affect the information matrix, and if so, how?

   Answer: We now have, for some symmetric positive-semidefinite matrix A:

   $$\begin{pmatrix} A & \mathbf{sparse} \\ \mathbf{sparse}^{\mathbf{T}} & I \end{pmatrix} \tag{2}$$

4. Say the robots cannot communicate for a while, and then communication is re-established. Will this affect the posterior estimate (at the time the communication is available again)? Argue your answer.

   Answer: No, it will not. The nice thing about the joint information matrix is that it is additive, and addition is commutative, so the order in which the elemts are added doesn't matter.