EL2320 Applied Estimation Lab2

Huijie Wang, huijiew@kth.se

November 2016

PART I - Preparitory Questions

Particle Filters

1. **What are the particles of the particle filter?**

   Particles are the samples drawn from a posterior distribution $bel(x_t)$, denoted $\chi := {x_t}^{[1]}, {x_t}^{[2]}, ..., {x_t}^{[M]}$.

2. **What are importance weights, target distribution, and proposal distribution and what is the relation between them?**

   Importance weight is the probability of the measurement z_t given particle ${x_t}^{[m]}$, i.e., ${w_t}^{[m]} = p(z_t|{x_t}^{[m]})$. They are non-normalized probability mass of each particle. Target distribution is a probability distribution corresponding to belief $bel(x_t)$, denoted f. And proposal distribution is a probability distribution corresponding to $\overline{bel}(x_t) = p(x_t|u_t, x_{t-1})bel(x_{t-1})$, denoted g.

   Since in particle filter it is impossible to drawn particles from target distribution f directly, instead we drawn particles from proposal distribution g. And importance weight is used to adjust g so that the distribution approaches f. Mathematically they are related by $w^{[m]} = \frac{f(x^{[m]})}{g(x^{[m]})} = \eta p(z_t|x_t)$.

3. **What is the cause of particle deprivation and what is the danger?**

   The problem occurs when the size of particles is not large enough to cover all relevant regions when preforming estimation. It happens especially in a high-dimension space. But it also could happen due to random resampling regardless of the size of particles.

4. **Why do we resample instead of simply maintaining a weight for each particle always.**

   Through resampling, the particles are forced back to target function and concentrate to high posterior probability regions. Maintaining a weight for each particle is still able to approximate the posterior, but the sampled particles will turn out to be spread in regions with low posterior probability. In this case, much more particles are needed in order to produce an appropriate representation of distribution.

5. **Give some examples of the situations which the average of the particle set is not a good representation of the particle set.**

   Poor estimation may occur when the distribution is a mixture model. In this case, the distribution can have several peaks and the average of particle set will fail to concentrate to several true states and group into one between the peaks.

6. **How can we make inferences about states that lie between particles.**

   Density extraction methods should be applied. We can make use of the particles within a certain distance with the target state and calculate the average. Also we can use histogram method which calculate particles in each bin or apply Gaussian Kernel to each particle.

7. **How can sample variance cause problems and what are two remedies?**

   Due to the randomness of resampling phase, the variance in the particle population decreases, in other words, the particle set loses its diversity. Therefore, the error of approximating its true belief increases.

   The two remedies are:

(a) Reduce the frequency of resampling.

(b) Use *low variance sampling*. That is, introduce a sequential stochastic process during the resampling process.

8. **For robot localization for a given quality of posterior approximation, how are the pose uncertainty (spread of the true posteriori) and number of particles we chose to use related.**

   With a larger uncertainty of pose, the number of particles should increase accordingly, in order to have enough particles covering state space. And vice versa.

PART II - Matlab Exercises

2 Warm up problem with the Particle Filter

1. **What are the advantages/drawbacks of using (6) compared to (8)? Motivate.**

   In (6) we only change x and y but remain *theta* unchanged. That means *theta* is only based on its previous step. The model is simpler but also comes with the problem that we can not measure the noise in angle *theta*. As a result, using (8) will make a smoother estimation based on noised angle in previous step.

2. **What types of circular motions can we model using (9)? What are the limitations(what do we need to know/fix in advance)?**

   As velocity v_0 and angular velocity ω_0 are fixed, the movement should be a perfect circle, regardless of the effect of noises.

3. **What is the purpose of keeping the constant part in the denominator of (10)?**

   In (10), the likelihood is represented by a Gaussian function, so that the constant part is a normalization factor which ensures the probability distribution add up to 1.

4. **How many random numbers do you need to generate for the Multinomial re-sampling method? How many do you need for the Systematic re-sampling method?**

   In Multinomial Re-sampling method, a random number within the range of $[0, 1]$ is generated for each sample, and M random numbers in total. On the contrast, in Systematic Re-sampling, only one random number is generated within the range of $[0, \frac{1}{M}]$.

5. **With what probability does a particle with weight $\omega = \frac{1}{M} + \epsilon$ survive the re-sampling step in each type of re-sampling (vanilla and systematic)? What is this probability for a particle with $0 \leq \omega < \frac{1}{M}$? What does this tell you? (Hint: it is easier to reason about the probability of not surviving, that is M failed binary selections for vanilla, and then subtract that amount from 1.0 to find the probability of surviving.**

   In multinomial resampling case, as a particle's weight is $\omega = \frac{1}{M} + \epsilon$, the range of the particle in CDF can be represented as $[\omega, \omega + \frac{1}{M} + \epsilon]$. As a result, if a generate random number does not lie in this range will lead to the particle not being resampled. For each resample step the probability of the particle not sampled is $1 - \frac{1}{M} - \epsilon$, and the probability of the particle not survived is $1 - \frac{1}{M} - \epsilon)^M$. So that the survival probability is $1 - (1 - \frac{1}{M} - \epsilon)^M$. Similarly, if the weight is ω, the survival probability is $1 - (1 - \omega)^M$.

   In systematic resampling case, as the cumulation length is $\frac{1}{M}$, so that any particle with weight no less than $\frac{1}{M}$ will survive, i.e. the probability is 1. On the other hand, if the weight is $0 \leq \omega < \frac{1}{M}$, the probability should be $\omega \div (\frac{1}{M}) = \omega M$.

6. **Which variables model the measurement noise/process noise models?**

   In the function, the measurement noise is represented by $params.Sigma\_Q$ and process noise is represented by $params.Sigma\_R$.

7. **What happens when you do not perform the diffusion step? (You can set the process noise to 0)**

   If we do not perform the diffusion step, all the particles will converge to the one with highest weight. Therefore, the sampled particles lose their diversity.

8. **What happens when you do not re-sample? (set RESAMPLE MODE=0)**

   Without resampling, the sampled particles can not be forced to the true target function but remains the random set when we initialize them. The particles turn out to be spread in regions with low posterior probability.

9. **What happens when you increase/decrease the standard deviations(diagonal elements of the covariance matrix) of the observation noise model? (try values between 0.0001 and 10000)**

   A smaller standard deviations of the observation model will decrease the variance of particles around the state, because it means that the measurement accuracy is more reliable. However, a too small covariance matrix will lead to the measurements not converging to the true state. That is because the covariance is too small so that the initial random samples hardly lie in the high probability range.

   In contrast, increasing the standard deviations of the observation make the particles converge around the true state with a large variance.

10. **What happens when you increase/decrease the standard deviations(diagonal elements of the covariance matrix) of the process noise model? (try values between 0.0001 and 10000)**

    The standard deviations of the process noise controls the variance of particles sampled. When decreasing, the diversity of particles also decreases. And the diversity of particles increases when increasing.

11. **How does the choice of the motion model affect a reasonable choice of process noise model?**

    The choice of precess noise should match the motion model and ensure the variance of particles are large enough to cover the true state. If not, the particles can not be forced to target function during resampling step.

12. **How does the choice of the motion model affect the precision/accuracy of the results? How does it change the number of particles you need?**

    The accuracy of the motion model directly effect the precision of result. If the motion model is not accurate and with large noise, the estimation will also be poorer so that more particles are needed to ensure the true state is covered within the particle cloud. And vice versa.

13. **What do you think you can do to detect the outliers in third type of measurements? Hint: what happens to the likelihoods of the observation when it is far away from what the filter has predicted?**

    A threshold can be set to detect the outliers. In the third type of measurements, noise is white Gaussian with small variance, but about 50% outliers are included. Since the outliers are far away from true state, the likelihood represented by the particle cloud can be very small. Therefore, choose an appropriate threshold can help detecting the outliers.

14. **Using 1000 particles, what is the best precision you get for the second type of measurements of the object moving on the circle when modeling a fixed, a linear or a circular motion(using the best parameter setting)? How sensitive is the filter to the correct choice of the parameters for each type of motion?**

    The results are shown as the following table:

Model	Measurement Noise: Sigma_Q	Process Noise: Sigma_R	Estimation Error
Fixed	diag([400, 400])	diag([20, 20, 0.1])	11.4 +- 5.4
Linear	diag([300, 300])	diag([2, 2, 0.01])	7.4 +- 3.9
Circular	diag([300, 300])	diag([2, 2, 0.01])	7.3 +- 3.6

As can be seen, the circular motion performs best as it accurately describes the true motion, and so that the noises can also be set to low values. Linear model also performs well. But in the case of fixed model, the covariance matrices have to be set larger in order to have a better estimation since the model is not accurate. In the three cases, the circular model is the one least sensitive to the parameters and the fixed model is the most sensitive one.

3 Main problem: Monte Carlo Localization

15. **What parameters affect the mentioned outlier detection approach? What will be the result of the mentioned method if you model a very weak measurement noise $|Q| \to 0$?**

    The outlier detection is effected by measurement noise Q as well as the threshold we chose. As measurement noise $|Q| \to 0$, it means the uncertainty of measurement is very low and the likelihood will be a thin and high peak, which result that most of the measurements will lie in low probability and be detected as outliers.

16. **What happens to the weight of the particles if you do not detect outliers?**

    Outliers usually have extreme values and if we do not detect outliers, they will be marked as valid and the particles will also move toward the outliers. Then it will end up with a bad representation of target function.

Datasets

1. map_sym2.txt + so_sym2_nk:

The results are shown as Figure 1 and Figure 2. In Figure 1(a) and Figure 1(b), 1000 and 10000 particles are used respectively. The results seems to be fine in the figures, but in reality, the particles tends to converge to wrong landmarks as the map is perfectly symmetric. That leads to the fact that there are 4 valid hypotheses in the map as there are 4 landmarks. A difference between using 1000 and 10000 particles is that not enough particle is more likely to lead to particle deprivation, but with more particles the system takes more time step to converge, as shown in Figure 1(c) and Figure 1(d), so that it is more stable in this case. However, the total time takes only 47.419341s in the case of using 1000 particles comparing to 191.471676s in the case of using 10000 particles.

On the other hand, the results of changing the noises weaker (0.01 times of origin) and stronger (10 times as origin) are shown as Figure 2(a) and Figure 2(b). In the case of using weaker noise, the system trust much more on the measurement so that it almost converges immediately in the first few steps. In contrast, it hardly converges in the case of using stronger noises. As can be seen, the 4 hypotheses survives until the end of simulation.

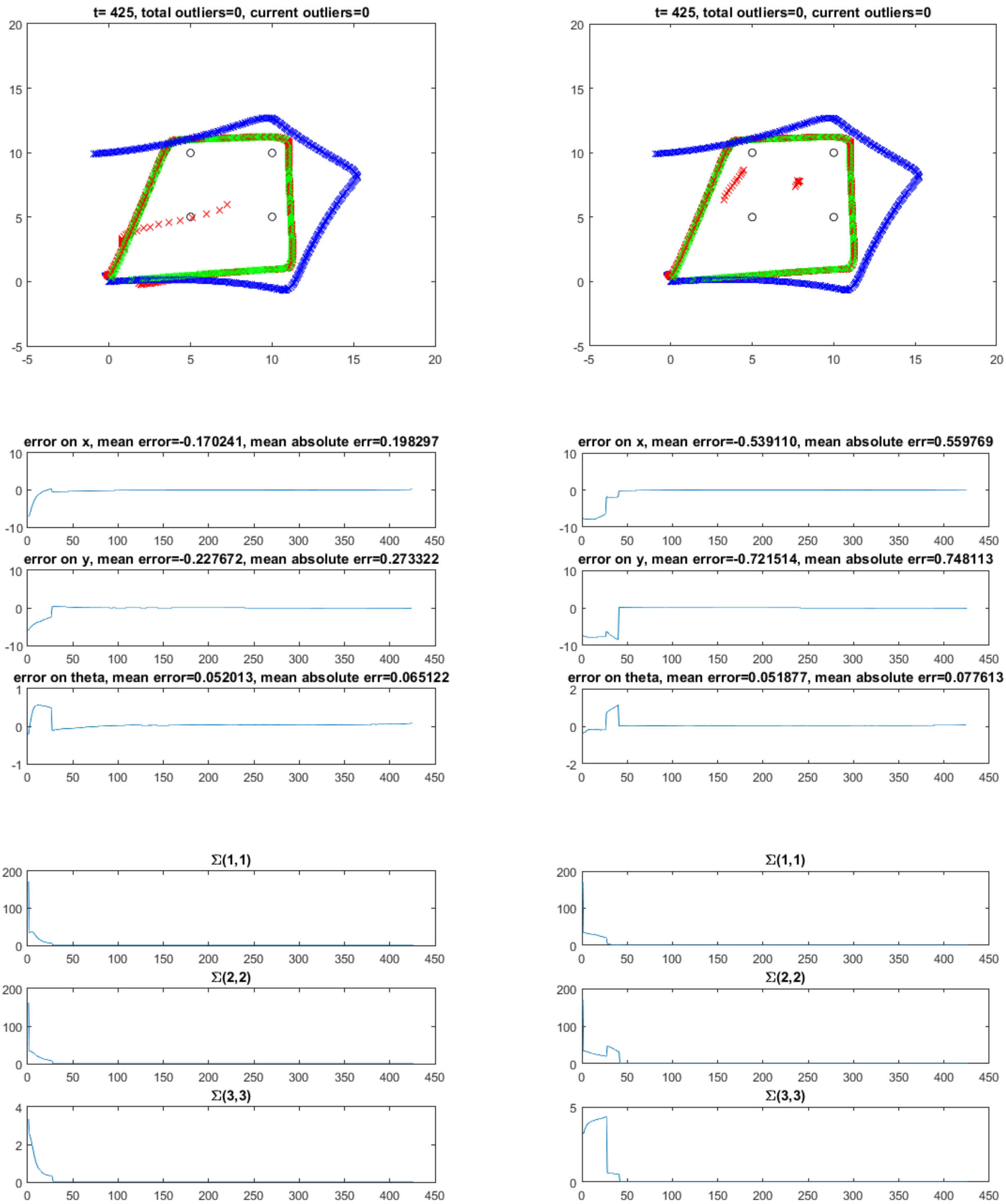

Figure 1: Result of using 1000 and 10000 samples

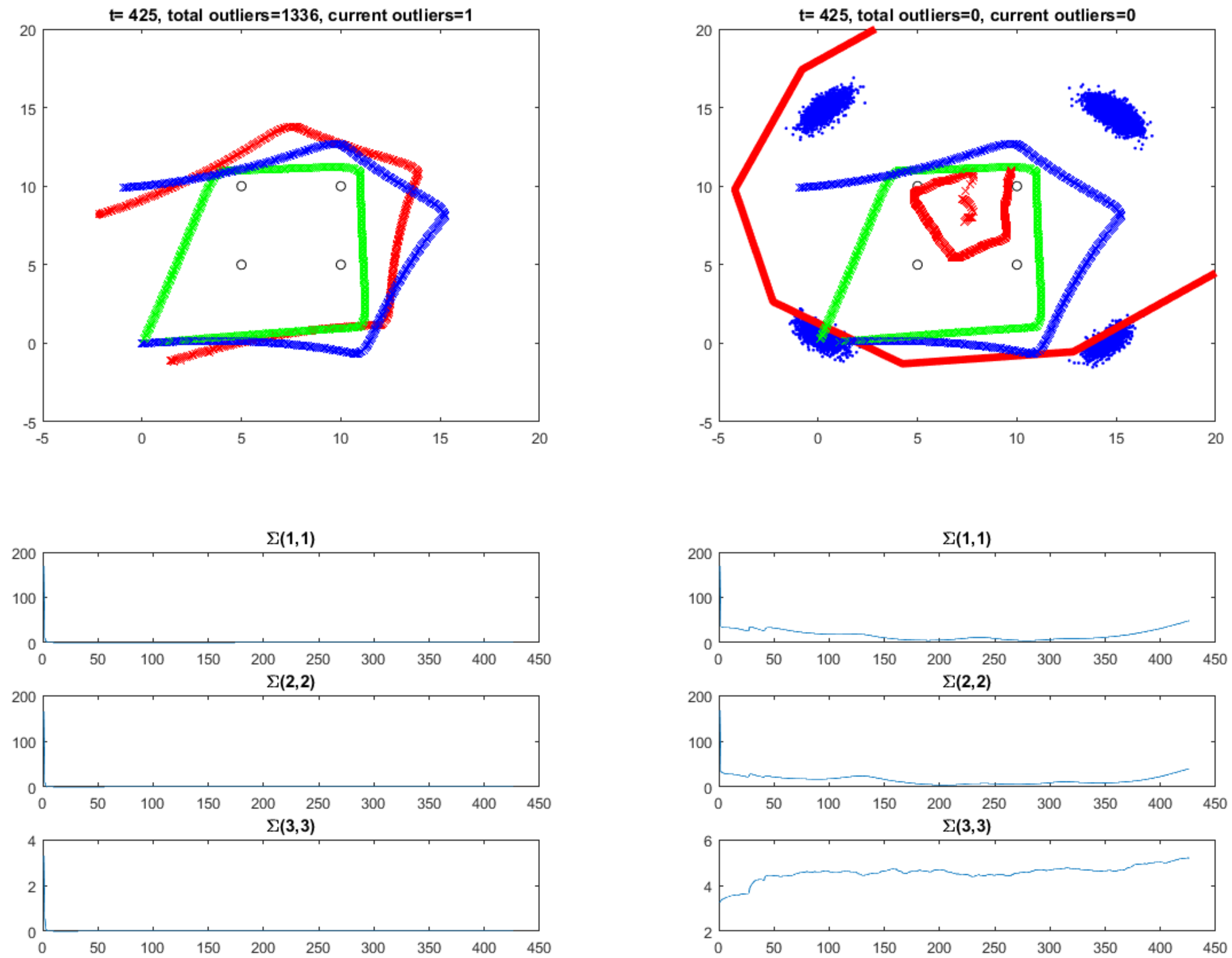

Figure 2: Result of changing to weaker/stronger noises

2. map_sym3.txt + so_sym3_nk:

The result can be seen in Figure 3. Here the noises are set to 10 times of origin so that the change of seeing the landmark breaks symmetry could be clearer. As can be seen, 2 hypotheses still survive until time step 181, but it quickly converge to the real one after seeing the fifth landmark.

In this case, it is important that the hypotheses still survive. If not, and unfortunately converge to a wrong hypothesis, the system can hardly recover even seeing the fifth landmark.

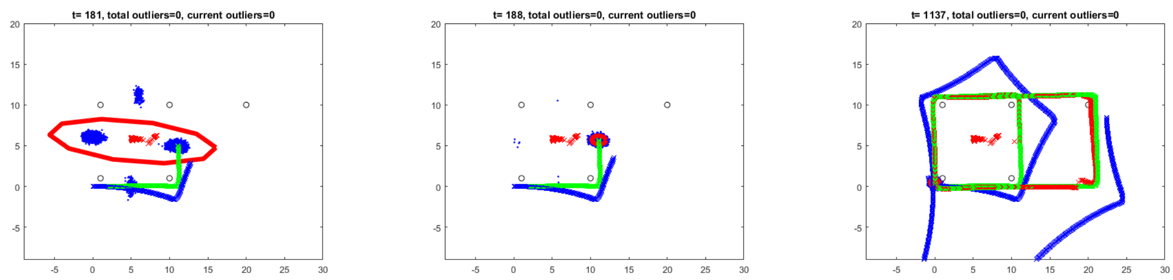

Figure 3: Convergence when landmarks breaks the symmetry