Student Name: Riccardo Periotto
Student ID: 19980915-T633

EL2320 - Applied Estimation
Lab 2: Particle Filter

Part I

Particle Filters

1. What are the particles of the particle filter

   In particle filters, a particle is a random state sample drawn from the posterior distribution over the state. Unlike the other filters (Bayes, Kalman and Extended Kalman), the particle filter is nonparametric, meaning that it represents the posterior $bel(x_t)$ by a set of particles, instead of using the parametric form.

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

   The importance weights are factors used by the filter to incorporate the measurement z_t into the particle set. Each weight is computed as the probability of the measurement z_t (passed to the filter) under the particle $x_t^{[m]}$ (to which the weight is associated). Mathematically:

   $$w_t^{[m]} = p(z_t|x_t^{[m]})$$

   The target distribution, denoted with f, corresponds to the belief $bel(x_t)$, while the proposal distribution represents the $\overline{bel}(x_t)$. This last distribution is denoted with g and is computed recursively by multiplying the model probability $p(x_t|u_t, x_{t-1})$ and the prior belief on the state $bel(x_{t-1})$. Distribution f is instead computed sampling from the proposal distribution, weighting each sample with its importance weight.

   The relation between the entities just described is summarized by the following equation:

   $$w_t^{[m]} = \frac{f(x^{[m]})}{g(x^{[m]}}$$

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

   The principal cause of particle deprivation is the variance introduced by random sampling. At each resampling step, the probability of wiping out particles close to the true state is larger than zero. For this reason, after a series of steps, it may happen that the number of samples covering all the relevant regions with high likelihood becomes too small. The danger is the lack of particles near the true state of the system.

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

   Without resampling, the filter would still approximate the posterior, but many of its particles would represent states with low posterior probability. With resampling instead, the filter refocuses the particle set to regions where the probability of the state is higher.

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

   One of the advantages of using a particle set is to be able to represent a distribution with multiple regions with high probability (i.e. multivariate normal distribution). However, with this approach, there are no guarantees that the average of the particle set is a good representation of the multiple regions of true interest.

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

   The continuos distribution over the state (that includes the states between particles) can be estimated through a procedure called density extraction. There are several ways to extract a density from a set of particles. Those we mentioned in class consist in:

   - fitting a unimodal Gaussian to the mean and variance of the particle set (as described in the previous question, this method may not be ideal in certain situations);
   - approximating the posterior with a multimodal distribution made by histograms;
   - using Gaussian kernels, in which a so called Gaussian kernel is placed around each particle ($p(x) \propto \sum_{m=1}^{M} w^{(m)} G(x - x^{(m)}, \sigma^2)$).

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

   The sample variance is given by the loss in diversity in the particle population. High sample variance can cause incorrect estimation of the true state and indeed the objective is to keep it as low as possible. There are two remedies to achieve this goal. The first is to reduce the resampling frequency, in particular when the state is known to be static. The second remedy is known as low variance sampling and it consists in using a sequential stochastic process for the state samples selection (instead of drawing them independently from each other as in the basic particle filter).

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.

   A higher pose uncertainty results in a larger spread of the posteriori distribution. To deal with this spread and being able to cover all the regions with probability, a higher number of particles needs to be used. On the other hand, a lower uncertainty results in less spread posteriori and consequently the need of particles is less.

Part II

Question 1

The advantages of using the 2D state space representation are the simplicity and the lower computational effort required. The angle is always constant and this gives the possibility to compute the value of the $\bar{u}_t$ signal only once at the beginning, using the found values as a constant afterwards. This approach works perfectly without the presence of noise.

On the other hand, the 3D state space representation models more complete and realistic behaviours. The noise affecting the system is considered through the adoption, in each time step, of the steering angle of the previous state. This representation also requires a larger number of particles with respect to the former.

Question 2

Given the constant angular velocity, the type of circular motion it is possible to model is the uniform one. The real system behaviour depends on the noise, but the uniform circular motion remains the underlying trend. The main law governing this motion is $v_0 = w_0 R$. As it is possible to see, one of the limitations is that we need to know two of the variables involved, while the other has to remain fixed. For example, we can assume to know the tangential velocity and the angular one, but then the radius of the circle the robot is following will be fixed.

Question 3

The purpose of the constant part is to normalise the value of the function. We want the function $p(z|x, \Sigma_Q, \bar{C})$ to be a probability function, and then the value of its integral over its domain has to be zero. The constant part is there to ensure this property. It is a factor common to all the Gaussian distributions.

Question 4

In the Multinomial Re-Sampling algorithm, the random generation is inside the cycle, so it is performed M times. In the Systematic Re-Sampling algorithm instead, the random generation is performed only once.

Question 5

With Multinomial Re-Sampling, the probability to select a particle corresponds to its weight $w^{[k]}$, so the probability of not selecting it is the complementary $1 - w^{[k]}$. The probability of not selecting a particle M times (the number of times the resampling is repeated) is then $(1 - w^{[k]})^M$. The complementary probability of this last event corresponds to the requested probability:

$$p(survival\_of\_particle\_k) = 1 - (1 - w^{[k]})^M$$

This corresponds to $1 - (1 - (\frac{1}{M} + \epsilon))^M$ when the importance is the one proposed in the question. With Systematic Resampling, a particle $p^{[k]} with importance weight w^{[k]} = \frac{1}{M} + \epsilon$ is always selected because the selection step is smaller than its probability. Here I am assuming that ϵ corresponds to a low positive value.

If the importance weight is $w_k \in (0, 1/M]$, the probability of selecting a particle with Multinomial Re-Sampling follows the same reason above, obtaining $1-(1-\frac{1}{M})^M$. In Systematic Re-sampling

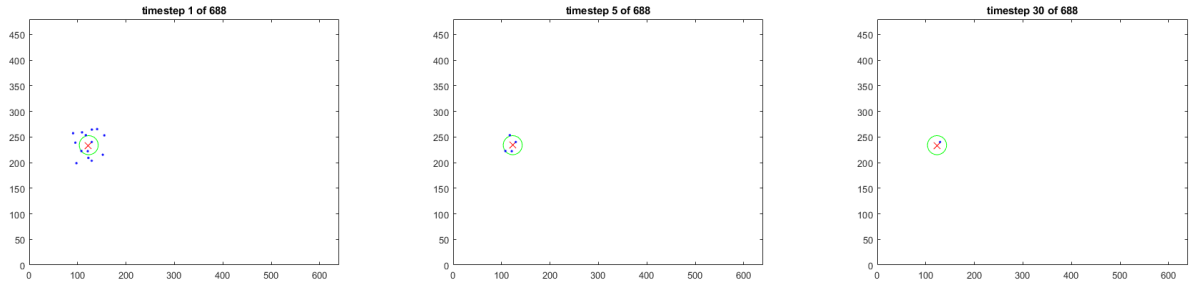
instead, the probability of selecting a particle with $w^{[k]}$ lower than $\frac{1}{M}$ is proportional to the weight itself and it is equal to $p(survival\_of\_particle\_k) = \frac{w^{[k]}}{\frac{1}{M}} = w^{[k]}M$ This value is always lower than one, meaning that there are no guarantees that the particle gets selected. Intuitively, the reason why the particle may be discarded is simply thatstep of the selection procedure skip it.

Question 6

Σ_Q models the measurement noise while Σ_R models the process noise.

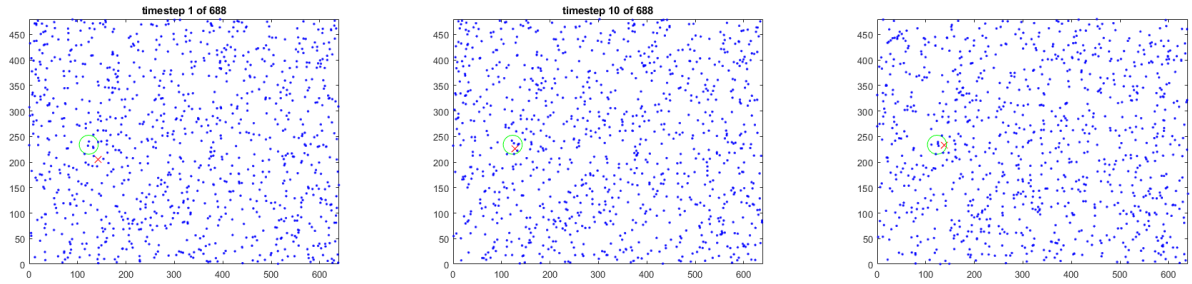
Question 7

Without the diffusion step, M identical particles will survive after a few iterations of the algorithm. It is impossible to know a priori after how many iterations this will happen, but it will happen for sure ("with probability one" as the book states). The diversity will disappear and the filter will believe that it knows the state of the robot, even if its belief will be wrong. Running the simulation, it is possible to see how after a few iterations (5 in this case), only two state samples are considered by the filter. After 30 only one sample is taken into account. Note that, in any case, the number of particles is always the same, M. They just collapsed in the same state sample(s).



Question 8

The resampling phase aims at transforming the distribution according to the particle weights. Without resampling, each particle simulates the estimation of the motion starting from their initial state (which, in this case, as we are considering the fixed simulation, corresponds to staying still). Independently of the number of iterations performed, the particles will be always spread over the whole state domain and the estimation error will consequently be large. The following figures represent the particle distribution after 1, 10 and 100 iterations. It is possible to see that there is no sign of convergence in the estimation and the only movement affecting each particle is only related to the noises (both in the process and in the measurement).

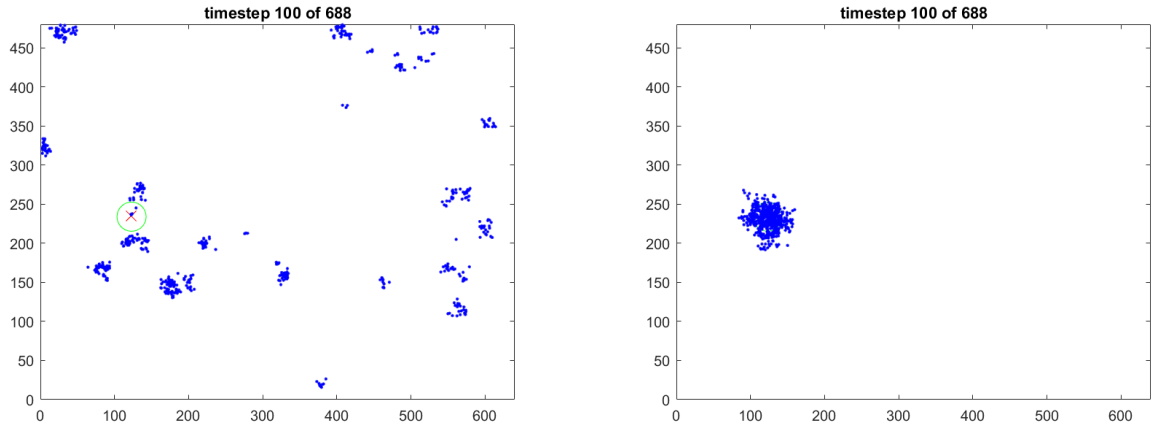


Question 9

A low measurements covariance makes the filter classify many of the measurements as outliers. In the code, to deal with the ouliers, all the weights are set to the same value $\frac{1}{M}$. Performing the resampling with this setting makes the filter to randomly select M particles from the set fed to it. The probability of selecting the same particles multiple times makes them cluster iteration by iteration. The more each cluster grows, the higher the probability to sample a particle from it. This makes the algorithm converge into a single cluster which, however, has no correlation with the real state.

A higher measurement variance solves this problem because makes the filter to be less strict with the outliers. The shape of the distribution of the measurements given the state is more spread and so the probability of selecting a particle close to the real state is higher. In a real application, this setting is preferred to the previous one, because it lets the filter to converge. The problem arises when the variance is too high. In this last case, all the particles are around the real state, but they are spread proportionally to the variance, possibly making the overall error large.

The two figures represent the situation of the particles after 100 iterations in the case of 0.1 and 100 variances.

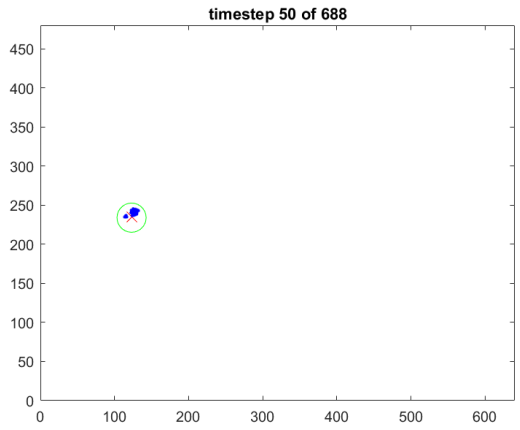


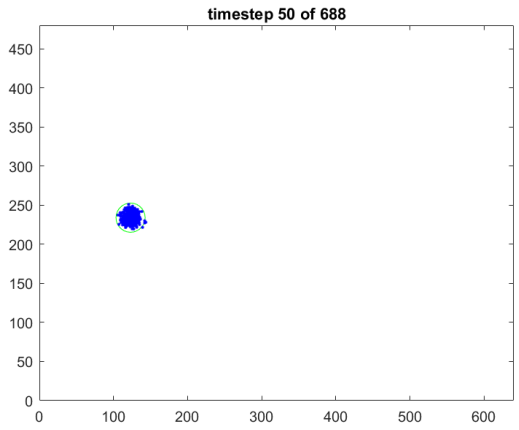
Question 10

This case resembles the scenario of Question 7. A low process covariance decreases the diversity in the particle population. This is connected to the fact that the covariance of the diffusion procedure is low and that the particles sampled by the filter can't move far away from the state prediction by the motion law. The particles tend to collapse into a single state sample, but there are no guarantees that this is the real one.

With high process covariance, the probability that the filter selects particles close to the true state is higher. The filter converges quicker, but the particles are around the real state following the process covariance (the higher the covariance, the larger the radius the more spread are the particles around the true state).

Figure on the left derives from running the simulation with all the diagonal elements equal to 0.1. The figure on the right instead derives from is the result obtained running the simulation with diagonal elements equal to 10.





Question 11

If the motion model is accurate, the process noise can be small, as the filter estimation is confident and resembles the real state. On the other hand, if the motion model is inaccurate, then the process noise should be larger, to let the particles estimate the true motion.

Question 12

If the motion model is accurate, the precision of the results is consequently higher. Connecting this with Question 7 and Question 10, it is easy to understand how, with an accurate model, a lower number of particles is required. On the other hand, if the model is inaccurate, its results are less precise and the number of particles required is higher. In any case, if the state space is modelled with three variables instead of two, the precision will likely again increase, together with the need for additional particles.

Question 13

A solution to detect the outliers could be the one that also proposed code implement. The way to identify it is to set a threshold on the average likelihood of getting the measurement considering the entire set of particles. If the average likelihood is below the threshold, then the measurement is discarted.

Question 14

The filter is especially sensitive when using the fixed motion type. In this case, the process and covariance matrices must be large enough to let the particles follow the actual state. How large depends on their combination. Simulating with small covariances and fixed motion type makes the particles stop in a certain region of the state space from which they are no more able to move away.

The linear and circular motion are less sensitive to this issue because, in addition to the diffusion, their predicted state is also given by a non-null motion model. For modelling the circular motion proposed in this example, the circular motion type is the one that results in the highest precision.

The best precision obtained when tracking an object moving on a circle with the three types is reported in Table 1, together with the values of the used parameters.

Motion type	Sigma_Q	Sigma_R	Error
fixed	diag([1000 1000])	diag([100 100 0.2])	11.1 ± 5.7
linear	diag([200 200])	diag([1.5 1.5 0.01])	7.5 ± 3.7
circular	diag([200 200])	diag([2 2 0.005])	7.1 ± 3.6

Table 1: Covariances and errors for the best precisions obtained with different motion types

Part III

Question 15

The two parameters affecting the outliers detection are the measurement noise Q and threshold λ_ψ.

As $|Q|$ gets smaller, the confidence of the measurements increases and the tolerance of discrepancies between the prediction and the observation values becomes lower. The result is a lower likelihood average that finds more difficulties in exceeding the threshold.

Following a similar reasoning, when setting a higher threshold, the average likelihood has to be higher for not considering the observation as an outlier.

Question 16

An outlier is a wrong measurement that the filter receives with a certain confidence. If the filter considers an outlier valid, it will resize its particle weights accordingly, even if the information provided by the outlier measurement is wrong. This step, instead of improving the estimate, makes it worse.

Dataset 4

With this dataset, the filter performs far better in the tracking task with respect to the localization one. The difference in performance with the standard parameters can be see comparing the figures in the left column and those in the right column in Table 2.

With this dataset, the filter performs far better in the tracking task than in the localization one. The difference in performance with the standard parameters can be seen by comparing the figures in the left column and those in the right column in Table 2.

The behaviour is aleatoric and the result changes simulation from simulation. The hypothesis deriving from the map symmetries are four, one for each landmark. However, as just said, what the robot follows is not deterministic. The general rule is that by increasing the number of particles, the filter can keep track of more hypotheses and it for a longer time. With a low number of particles, the system is subject to the phenomena considered throughout the assignment called "particle deprivation". Together with the number of particles, also the measurement noise plays a fundamental role in how the system behaves. Setting a high noise causes the system to converge slowly while leaving a low Q let the estimation converge faster, but not necessarily in the right state.

The observed difference between the two sampling methods is that the estimation converges faster in one of the hypotheses, but as it doesn't know which one is correct this is not the behaviour we would like to observe. Hence systematic resampling is better.

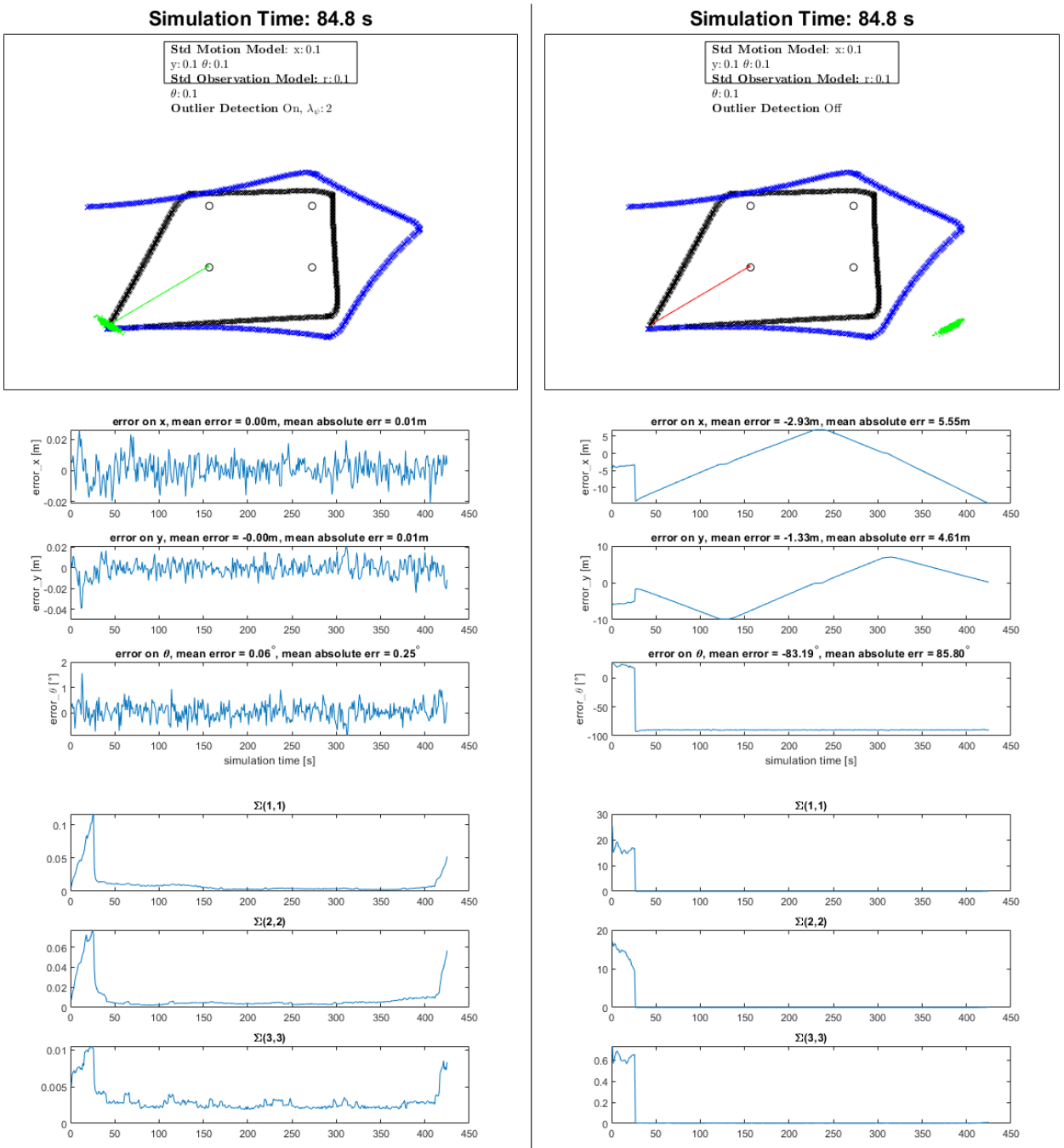
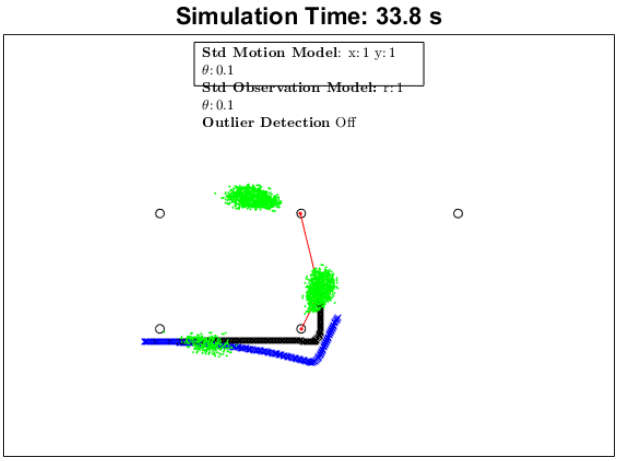
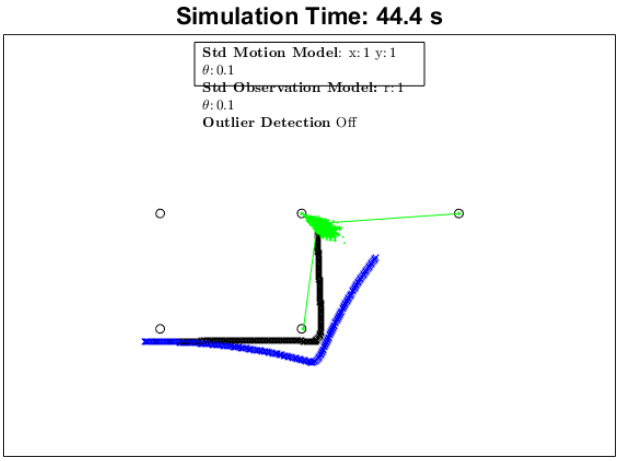


Table 2: comparison between the tracking and localization tasks with dataset 4

Dataset 5

The two images requested are reported in the following figures. I had to increase both the noise and the number of particles to let the system track the different hypotheses.


Simulation Time: 33.8 s
Std Motion Model: x: 1 y: 1
θ: 0.1
Std Observation Model: r: 1
θ: 0.1
Outlier Detection Off


Simulation Time: 44.4 s
Std Motion Model: x: 1 y: 1
θ: 0.1
Std Observation Model: r: 1
θ: 0.1
Outlier Detection Off