We would like to thank you and appreciate your clear review and invested time. Before going through your comments, we want to inform you that we have added the results of another high dimensional experiment, visual odometry task. Please note that these results were almost ready at the time of the submission, but due to the page limit we decided to cover two high dimensional experiments, single-double pendulum, and two low dimensional experiments, Lorenz and NCLT. But, because of the concerns of some reviewers about higher dimensional experiment results, we have merged single-double pendulum experiment and added visual odometry results in the rebuttal version.

We try to clarify point by point based on your comments and feedback.

The notations are a bit confusing. For example, (1) the the encoder function $f(.)$ and decoder function $h(.)$ do not have connection with the state transition and emission matrices Fs and Hs. (2) Comparing Figure 1 with Figure 2, the output of encoder and the input of decoder seem to be not consistent in the sense that $w : (w, r)$; $x : (x, Sigma)$. We know r and Sigma are uncertainty variables but you may better show that without confusion.

R1.

1- Please note that the encoder $f(.)$ constructs latent observations w from higher dimensional noisy observations o and decoder $h(.)$ reconstruct high dimensional noise free posteriors o^+ from posterior state (x^+, Σ^+). If the length between time-steps of observations is sufficiently small, then the constructed latent model is linear or can be closely approximated by a good linear approximator, which is the case in our experiments, so F and H can be sufficient to describe dynamics in the latent space. Thus we use dynamic parameters, H and F, in the transition block (fig1) to obtain next posterior (Σ^+, x^+)(decoder input) from current received observation w(encoder output). Accordingly, we just need (H, F) in the transition block (fig1) and there is no need to connect them to enc/dec.

2- Thanks for your recommendation about fig1. We have modified it in the rebuttal version. Please let us know if you still find it confusing.

To deal with high-low dimensional observations, is the core idea coming from the how do we model the nonlinear function $f(.)$ and $h(.)$?

R2.

Non-linear functions, $f(.)$ and $h(.)$ in the paper handle the non-linearity of the system, meaning that more complex $f(.)$ and $h(.)$ could handle more non-linearity and linear assumption of transition block may be more realistic. We can divide the core idea of paper into two parts:

1- The abilities of $f(.)$ and $h(.)$ for constructing noise-free observations w and reconstructing high dimensional noise free posteriors o^+, such that the transition block be linear(or at least be approximated linearly)

2- The abilities of the transition block for (i) performing an accurate linearization of the system, where it uses more complex function approximators like GRU cell and dynamic networks and

(ii) conducting filtering/smoothing to obtain posteriors.

By having these two, we are able to handle high-low dimensional with noisy/missing observations with/without known dynamics.

As for the dec(cover), how do you model the correlation access different dimensions?

R3.

dec(covar) includes a DNN with the input of $\mathbf{\Sigma}^+$ and output of a lower triangular matrix L with positive diagonal elements, then calculate dec(covar) = LL^T. By this solution, Cholesky factor consideration, we ensure both the positive definiteness and determine the correlation and non-diagonal elements of dec(covar).

Figure2 looks too complicated ...

R4.

To address your concern, we have proposed three figures that fig2 can be replaced by them:

1- An overall structure 1, without subfigures in this link

2- An overall structure 2, without subfigures in this link

3- An overall structure with three separated subfigures in this link

We like to have your opinion about them.

We hope that we have been able to address the concerns about fig2 and organizations and will be happy to answer any further questions or any potential modifications about the organizations.