Review on manuscript
"Bayesian Learning of Degenerate Linear Gaussian State Space Models using Markov Chain Monte Carlo"
by P. Bunch, J. Murphy and S. Godsill

In this paper, the authors address the Bayesian learning of *degenerate* linear and Gaussian state space models. In this specific context, classical Gibbs sampler can not be used and therefore the authors propose an interesting and complete framework based on Markov Chain Monte Carlo (MCMC) to deal with this challenging problem. The paper is well written and organized.

I have the following remarks:

- Even if I found the paper well organized, I think that the authors should summarize the complete framework (RJ-MCMC) in an algorithm - in my opinion it will help the reader to understand it by summarizing all the steps of the proposed scheme in one place.
- In Section III-C, the authors propose to use a collapsed Gibbs move to more efficiently exploit the target on the space of F and Q, alternatively. I think the authors should discuss in more details why such collapsed Gibbs is important in the text. Moreover, the importance of such steps could be easily illustrated in the results section so the reader can appreciate the benefit of using such a strategy.
- My main concern is regarding the simulation section.
  - Firstly, the comments are just a brief summary of results from the figures without any comments or discussions on why such results are obtained.
  - There are no figures that show the evolution of the Markov chain OR the posterior approximation of the unknown rank r - The authors wrote that *"the sampler identifies the true rank.... and remains at this value thereafter"* but this can be worrying - this statement could also mean that the sampler is not mixing properly for this variable r.
  - The author should be more precised by saying *"Trace plots suggest that the sampler converges within a few thousand iterations. See figure 2"* - The convergence showed in this figure is only related to the variable ξ_y
  - the authors could illustrate in the first toy example the benefit of the collapsed Gibbs moves by showing the performance of the sampler with vs without this strategy.
  - I personally think that the proposed algorithm should be illustrated with at least a different set of parameters for the toy example (e.g. different values for ξ_y thus allowing to have a curve of RMSE vs ξ_y)

List of minor remarks/typos:

- Just after Eq. (16), the authors should be more precise on the following sentence to avoid any confusion: *"However, whereas before Q was required to be positive definite, this is no longer the case"* (I would add) *for unknown G.*
- Please correct in Section V first sentence *... known a prior.* by *... known a priori.*