We thank the reviewers for their thorough review and insightful feedback and appreciate their positive comments.

Experiments: We agree with the reviewers that further empirical evaluation of our algorithm spectral Kalman filtering (SKF) is beneficial. We will add more experiments to the final manuscript. Here we present our preliminary work on these experiments. Minor details on experiments are excluded due to space constraint.

(1) *Hyperparameter sensitivity:* The hyperparameters of SKF are the number of filters k and the regularization parameter α. In the paper, we set $k = 20$ and use $\alpha > 0$ (in particular, $\alpha = 10^{-4}$) when the empirical feature covariance is singular, which we observe only happens in the first two time steps. Per Reviewer 2's suggestion, we analyzed the hyperparameter sensitivity of our algorithm. Figure 1(a) shows that our algorithm is robust with respect to hyperparameter k.

(2) *Perturbation analysis:* Per Reviewer 1's suggestion, we analyze the robustness of SKF in the presence of nonlinear perturbation according to logistic growth dynamics $h_{t+1} = Ah_t - \beta h_t^2 + Bx_t + \eta_t$. We vary β and choose x_t such that observations do not explode. Figure 1(b) illustrates average prediction error vs. β that shows the performance of our algorithm degrades gracefully with non-linear perturbation.

(3) *Comparison with system identification (SI) methods:* Per Reviewer 3's suggestion, we plan to compare our algorithm with SI-based methods for the final paper. Here we present one experiment comparing the performance with EM followed by Kalman filtering (Figure 1(c)). In addition, we showed superior performance over wave filtering of Hazan et al. (2017), who demonstrated that their algorithm works better than SI-based methods. SI-based methods, besides being often significantly slower, either do not have regret guarantees or have degrading performance when $\rho(A) \to 1$.

Reviewer 1: [*Non-symmetric* A] We compare the algorithms in both systems with symmetric and non-symmetric A. Although the theoretical analysis of wave filtering is given for symmetric A, the authors mention in their paper that empirically their algorithm also handles non-symmetric A. System 3 experiment compares the performance of wave filtering when A is non-symmetric and $\rho(A) = 1$, which results in a growing observation norm. [*System parameters*] For System 2, we set the parameters somewhat arbitrary to encounter a system that exhibits long memory. While the spectral norm of A and G are parameters of interest that distinguish the performance of algorithms, we were not concerned with particular values of other parameters. [*Perturbation analysis*] We thank the reviewer for their suggestion. We have included an experiment here and will include further analysis to the updated manuscript. [*Wave filter implementation*] While we were not provided with the source code, Hazan et al. (2017) gives a detailed explanation of their algorithm in their paper and appendix. We implemented wave filter with follow the leader algorithm (as opposed to projected gradient descent) as suggested in their paper for better performance and by our observations that tuning the step size for gradient descent implementation of wave filter was very difficult. [*Source code*] We plan to make the code public after optimizing the implementations. [*Clarity*] We agree with the reviewer that the clarity of the approximation theory section can be improved. We will draw connections with Temlyakov's results and elaborate more when updating the manuscript to make it as accessible as possible. [*Citations*] We thank the reviewer for recommending additional references. We will include them in the final paper.

Reviewer 2: [*Real-world experiments*] We agree with the reviewer that testing our algorithm on real applications is important. We are collaborating with a team to apply the proposed algorithm to healthcare where long-term forecast memory is critical. [*Hyperparameters*] We have provided the hyperparameter k in the paper and have discussed α above. We agree that analyzing hyperparameter sensitivity is critical for practical considerations and will be included.

Reviewer 3: [*Experiments*] We thank the reviewer for their suggestion and refer to the above segment on experiments. [*Comparison with Hazan et al. (2017)*] Our algorithm shares some similarities with wave filter but has important differences: (i) We design our filters by applying spectral methods to Kalman predictions which results in different filters compared to those in wave filter (ours is related to the covariance of $[1, \lambda, \ldots, \lambda^T]$ for $\lambda \in [-1, 1]$ while theirs is related to $[1 - \lambda, \ldots, \lambda^{T-1} - \lambda^T]$ for $\lambda \in [0, 1]$). This filter construction poses many theoretical challenges such as growing feature norm and Lipschitz constant but is crucial for convergence to Kalman prediction which was left open in Hazan et al. (2017). (ii) Wave filter only includes the most recent observation but our algorithm considers all past observations. (iii) Wave filter is based on projected gradient descent whereas our algorithm uses regularized least squares. [*Algorithm name*] We concur with the reviewer and consider a more descriptive name (such as Spectral LDS Adaptive Predictor). We are also working on a recursive version and hope to achieve that for the final version.

Reviewer 5: [*Control inputs*] While the algorithm derivation, improper learning approximation error, and most of the regret analysis consider control inputs, the excitation result is given without inputs. We believe that extending our analysis for LDS with inputs is possible by characterizing input features and in light of the experiments. As pointed out by the reviewer, such an extension requires some care. For instance, one needs to characterize the covariance between features constructed from observations and features constructed from inputs. We will add a discussion to the final paper. [*Notation and citation*] We thank the reviewer for their suggestions and will incorporate them in the final paper.

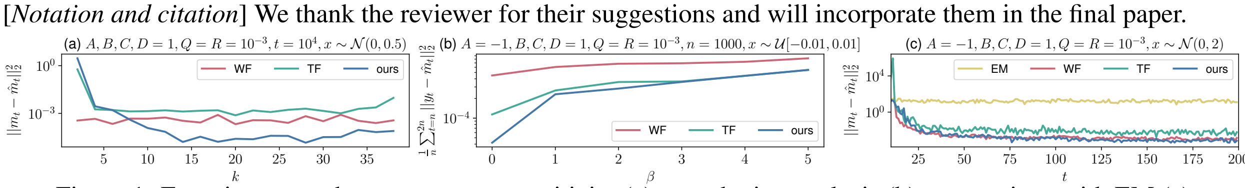

Figure 1: Experiments on hyperparameter sensitivity (a), pertubation analysis (b), comparison with EM (c).