|  |  |  |
| --- | --- | --- |
| 1 | Author Feedback | **Novelty of the proposed model: Most previous work on kernel embedding of distributions, such as two sample tests and independence tests, do not consider the presence of latent variables. We consider novel aspect of the spectral properties of kernel embeddings and connect them to the spectral properties of infinite dimensional tensors and provide a sample complexity analysis.   We are not aware of any prior work using kernel embeddings to explicitly recover the parameters of the latent variable model. The related work of Kernel embedding of hidden Markov models (Song et al. ICML 2010), and Kernel embedding of latent tree models (Song et al. NIPS 2011) focus on estimating the marginal density of the observed various but not the latent parameters. In contrast, our algorithm can recover the exact latent parameters, and we prove that the estimator is consistent and analyze its finite sample complexity in this work.   Novelty of technical analysis: In terms of the analysis, previous work on tensor decomposition (Anandkumar Et al 2012) provided analysis when there is perturbation on the orthogonal tensor. In our current paper, we take into account the whitening perturbation and the sample bounds to achieve a certain level of concentration of the empirical moments and present a unified sample complexity analysis. This is novel.   Comparison to other methods in experiments: We have compared our method with the best known algorithms for estimating latent variable models (EM), tensor approach for Gaussian mixtures (Hsu and Kakade) and the recent nonparametric algorithm of Kasahara & Shimotsu, 2010. We are not aware of any other nonparametric methods for recovering parameters of latent variable models. Indeed, EM and Hsu&Kakade approaches fit to a limited class of models since they are parametric. On the other hand, they are fast to run and are popular approaches for learning latent variable models. Thus, we selected them for comparison. The approach of Kasahara and Shimotsu results in very poor results. Hence, we do not plot them (line 629 in main paper).  Experimental details: We've corrected the dataset number to 24 to be consistent. We \*only\* split the dimensions but not the datasets. The dimensions are split according to the correlation between dimensions, so that weakly correlated dimensions go into different views.** |
| 2 | Review Quality Feedback (Optional, Visible to Area Chairs Only). | **It is disappointing to receive only one review for the paper. We request additional expert reviews for the paper. We have now clarified the concerns of the reviewers concerning novelty and the experimental results. The reviewer incorrectly mentioned that we do not compare with other non parametric approaches. We compare our method with the work of Kasahara and Shimotsu in our experiments, which is a non-parametric approach, but did not plot them since they resulted in very poor performance. We mention this in text (line 629 in main paper).** |