October 22, 2016

Re: Resubmission of manuscript *An M-Estimator for Reduced-Rank System Identification*, PRLETTERS-D-16-00643

Dear Editors,

Thank you for the opportunity to revise our manuscript, *An M-Estimator for Reduced-Rank System Identification*. We appreciate the careful review and constructive suggestions. It is our belief that the manuscript is substantially improved after making the suggested edits.

Following this letter are the editor and reviewer comments with our responses in italics, including how and where the text was modified. The revision has been developed in consultation with all coauthors, and each author has given approval to the final form of this revision. The agreement form signed by each author remains valid.

Thanks for your consideration

Sincerely,

Shaojie Chen, PhD

**Editor:**

In this paper, the authors have proposed a method to address the problems of fitting statistical models to high-dimensional data. The paper was reviewed by two independent experts in this field. The reviewers have raised some concerns about the method, symbols and lack of comparisons. So, I would suggest the authors to prepare a revision by thoroughly addressing the comments of the two reviewers.

**Reviewer #1:**

The authors describe an algorithm for efficient estimation of parameters of linear dynamical systems from high dimensional dataset. The algorithm is a variation of the standard Expectation Maximization algorithm; the performance has been evaluated on the basis of a simulation; an example of application on fMRI data has been also presented.

1. The topic falls within the scope of the journal, English is sufficiently clear, although it might preferable a revision by an English native speaker.

*We have asked two more native speakers to revise the article and polish the English.*

2. It is not clear to me why the authors have emphasized the 'M-estimator' property, they could have used 'Maximum Likelihood' equally well. (See for example Wikipedia M-estimator)

*Thank you for this interesting comment. The M-estimator notation captures a discussion we have had over a long period of time prior to submission. The original name for the model actually used “maximum likelihood” instead. The motivation for the change to ‘M-estimator’ is that when we introduced the penalty term to the log-likelihood, the new objective function was not the original log-likelihood. The log-likelihood is related to a posterior with a particular prior. However, we found that connection not worth discussing further as the connection with Bayesian inference is formal only*

3. Introduction: second paragraph: it is said that high dimensional time series have: what is? Please define each quantity before its use.

*Thanks for the observation. Here refers to the dimension of the high-dimensional time series. We have rewritten the introduction to define before it is used.*

4. Section 2 , 'the model', is in general sufficiently clear, however, before equation (3) the letter 't' has been used to indicate time, after equation (3) the letter 't' has been used to represent a threshold this is confusing, please correct it.

*To avoid this, we now use the Greek letter to represent the threshold used in the dual form of problem (3).*

5. Section 3, the 4th paragraph ends with '(Boots, Boots)' what does it mean?

*This (Boots, Boots) was intended to cite the paper “Learning Stable Linear Dynamical Systems” by Byron Boots. Its strange format is caused by an incomplete BibTeX citation item, which is missing the publication year. We have fixed this issue by using the complete BibTeX citation item.*

6. Section 3.1: the formula for the expectation of the log-likelihood of the complete data seems to be wrong, the correct formula should be: E [log p(X, Y) | Y]; it seems the same error was done in the appendix.

*This is correct. The E-step of EM algorithm requires the conditional expectation of the log-likelihood, not the expectation of the conditional likelihood. We have made the correction throughout the article and Appendix.*

**Reviewer #2:**

The proposed method introduces penalty terms to the linear dynamical system (LDS) which is regarded as a continuous version of the hidden Markov model (HMM). The optimization problem is shown in Eq. (3) and an extended EM algorithm is proposed to estimate parameters of the proposed method.

1. Note that the proposed method still applies the statistically independent assumption on P (Y|X) as seen in the first equation in Section 2. This is still a big challenge in HMM and other methods for time series data and the proposed method have not addressed this one.

*This is accurate. The assumption of conditionally independent observations given latent states remains an unsolved challenge in HMM/LDS, especially for high dimensional time series. Nonetheless, conditionally independent models continue to prove useful in applications. We have not addressed this issue in this work. In the discussion, we briefly discuss recent work on this topic and our next steps.*

2. It is not clear why the introduction of penalty terms can improve the classification/prediction accuracy. Experiments present results for estimation accuracy versus penalty and a highest accuracy point is found. However they are the best parameter values for the proposed method. This does not mean that the proposed method would be better than other existing methods in the literature.

*The reviewer is absolutely correct, they are the best parameter values for the proposed method and this does not imply that the proposed method would be better than other existing methods. When the penalty applied to MR.SID is 0, the model is exactly equivalent to the generic LDS. For the neuroimaging application in this work, the built-in property of the data fits well with the sparsity assumption of MR.SID, so the algorithm gives better results than the generic LDS. For some other applications, with no inherent sparsity, the optimal penalty might be 0 and the algorithm should give identical result as the generic LDS/HMM.*

3. Experiments in the article only show comparison results with Singular value decomposition (SVD), however no experiments were performed to compare the proposed method with other existing methods, for example LDS and HMM.

*Thanks for the suggestion. Adding a comparison between the proposed algorithm and LDS/HMM is necessary as LDS/HMM is the baseline model that MR.SID is based on. Thus, we have included a comparison of the out-of-sample predictive power of MR.SID and LDS/HMM – the result is summarized in the third panel of Figure 4. MR.SID has better out-of-sample prediction accuracy than the generic LDS/HMM. In addition, the prediction accuracies of MR.SID with different penalty sizes are also plotted in the same figure – and we can see similar patterns that we saw from simulation (Figure 2).*

**Associate Editor:**

In view of the comments made, the Associate Editor responsible for your article has decided that the paper can be reconsidered for publication after revision. Therefore we look forward to receiving the revised version of the article as soon as possible, but before Dec 01, 2016 (unless the Editor specifies otherwise), together with a list of changes or a rebuttal against each point which is being raised.

If we do not receive your revision within this period, your paper will be considered withdrawn. Any revision after this will be treated as a new submission and subject to a new review cycle.