Response to comments

- Fixed typo and grammatical errors
- Added detailed plot description.
- Hid codes.

- Per-comment response
    - *Why not change the for-loop in algorithm 1 for a while loop and give a constraint? That way you will actually achieve a better estimate for the MLE.*
        - The goal is to find asymptotic distribution of a finite-time estimator $\hat{\theta}^{(T)}$. Arbitrary number of iterations resulting in while loop makes it not suitable. However, you can use while loop in practice to guarantee good convergence.
        - In addition, if we set the maximum number of iteration T as the minimum that satisfies convergence criteria, the for loop can be viewed as a generalized while loop in this case.
    - *I see no reason to write $\hat{s} = d$ instead of $s = d$. In other words, why have an estimator of s?*
        - We do not know the true sparsity parameter s. If we want to perform high-dimensional sparse EM, estimation of s should be done beforehand.
        - For the rest of the blog, we covered non-sparse case ($s = \hat{s} = d$).
    - *You use λ in the definitions, but it is not clear where λ comes from.*
        - We already provided the detailed form of $S_n(\theta, \lambda)$. In this form, λ is used inside of the constraint of Dantzig selector problem.
    - *"ζ scales with d,n". What rate does the paper say ζ scales at with respect to n, d?*
        - The scaling rate differs from model to model. We specified an example of the form of ζ's for a Gaussian mixture model in the succeeding section.

We felt the need to comment more on reviewer's personal thoughts. Indented italic paragraph is a direct quote.

> *It is not clear to me why one might be interested in the asymptotic properties of an "EM estimator" since EM is a method for maximizing a likelihood or objective function. This would be the equivalent of investigating asymptotic properties of an "Newton's Method estimator", which does not make much sense as Newton's method is an optimization algorithm rather than a statistical method.*

First, ReML estimator of a regression model is a special form of "Newton's Method estimator". However, we still are interested in finding the uncertainty of it. In fact, major statistical software including SAS and R base package provides asymptotic standard error estimates using the fact that ReML estimator admits asymptotic normality/chi-square. The critical difference between ReML and Newton's method is that the objective function of ReML has random elements (due to a random sample) in it. This allows, and encourages, us to quantify the uncertainty of the resulting estimator.

Similarly, EM algorithm is indeed a special form of majorization-minimization (MM) algorithm, a well-studied optimization method. EM algorithm is a "special form" in the sense that it maximizes Q_n, which incorporates a random sample in it. There is nothing nonsense in investigating the statistical property of a random object.

> *EM algorithm is commonly used to find MLEs, so it would not be surprising if "EM estimators" had similar properties as MLEs. The more interesting part is using values derived from the Q function to estimate the score and information matrix.*

EM algorithm is indeed commonly used to find MLEs. It is especially heavily used for parameter estimation of latent variable models. A common problem is that the likelihood is not concave, and we cannot guarantee that the EM estimator is an MLE. The result of this paper is interesting because it showed that even a local optimum asymptotically follows normal distribution. In fact, we used notations $S(Y, \theta)$ and $\hat{I}$ to emphasize the link between Q_n function and the ordinary Wald and score test statistics. We wanted to make the Wald-type and score-type test statistics for EM look like that of MLE. The paper did not used such notations and did not emphasize the estimation of score and information matrix.