We greatly appreciate all the insightful comments and valuable suggestions on how to improve the paper.

One major concern raised by the referees centers around the use of nonconvex penalties. Here we would like to clarify our choice of nonconvex penalty over convex penalties. At a high level, the rationale for using the nonconvex penalties is that nonconvex penalties in high-dimensional models have been shown to outperform its convex counterpart both in theory and in practice in many situation [see, e.g., Fan and Li, 2001, Zhang, 2010, Shen et al., 2012, among others]. In the context of this work, we agree that convex penalties can also be used. However, the corresponding theoretical results may be less sharp in terms of both convergence rate and the conditions needed for sparsistency. Moreover, nonconvex penalty tends to perform better as compared to its convex version in practice. To empirically verify this in our context, we have performed additional simulations using convex penalties. The results are summarized in Table 1, in which "Our-ML" and "Our-ML-convex" denotes the proposed method using nonconvex and convex penalty, respectively; and "MRCE-group" denotes the MRCE method using the same convex penalty (with the old parameterization). Table 1 shows that the proposed methods with convex penalties performs much worse than the nonconvex version. More importantly, the advantages of the new reparameterization is evident by comparing our method using convex penalty ("Our-ML-convex") with MRCE using the same penalty ("MRCE-group") as they only differ in parameterization. Therefore, this means that using the proposed parameterization leads to better estimation accuracy. Finally, we would like to point out that the nonconvex penalty we used includes the convex L_1 penalty as a special case when $\tau = +\infty$.

Regarding to the proposed reparameterization, one referee asked about the novelty of the proposed reparameterization given that the negative loglikelihood function is convex in a classical reparameterization. Here we would like to clarify that the classical reparameterization $(B, \Theta) = (C\Omega^{1/2}, \Omega^{1/2})$ could be used if we do not want to impose sparsity-inducing penalty on Ω. As one referee have pointed out, one can still impose a penalty on B to retain the row-sparsity structure. However, it seems to be difficult to penalize Θ so that Ω is sparse. This is why we propose a different reparameterization so that we can impose sparsity-inducing penalties on both C and Ω.

Regarding to the computational details of the DC algorithm, we agree that this part was not clearly presented. We modified line 208-211 as follows

Its key idea is to decompose the objective function into difference of two convex functions, and linearize the trailing function to obtain an upper convex approximation of the nonconvex objective. In our setting, using the DC decomposition that $p_\tau(x) = |x| - \max(|x| - \tau, 0)$*, we obtain upper convex approximation of the nonconvex penalty at the previous iterate* $\left(B^{(t)}, \Omega^{(t)}\right)$*:*

$$p_{\tau_B}\left(\|B_{i\cdot}\|_2\right) \le \|B_{i\cdot}\|_2 \mathbb{I}(\|B_{i\cdot}^{(t)}\|_2 \le \tau_B) + \tau_B \mathbb{I}(\|B_{i\cdot}^{(t)}\|_2 > \tau_B)$$
$$p_{\tau_\Omega}\left(|\omega_{ij}|\right) \le |\omega_{ij}| \mathbb{I}(|\omega_{ij}^{(t)}| \le \tau_\Omega) + \tau_\Omega \mathbb{I}(|\omega_{ij}^{(t)}| > \tau_\Omega)\,.$$

We have created a new revision which fixes the typos found by the reviewer and makes some minor changes to the technical proofs to make them more accessible. We are glad the reviewer appreciates the main ideas behind our paper, and in the next revision of the paper, we plan to incorporate your suggestions and the above proposed changes.

Table 1: Additional simulation results.

(p, q)	Method	Error($\widehat{C}$)	Error($\widehat{\Omega}$)	Error($\widehat{C}, \widehat{\Omega}$)	FPR($\widehat{C}$)	FNR($\widehat{C}$)
$(200, 3)$	Our-ML	.051(.028)	.031(.019)	.085(.038)	0(.002)	0(0)
	Our-ML-convex	.371(.082)	.684(.073)	.933(.058)	.434(.062)	0(0)
	MRCE-group	.597(2.96)	.433(1.18)	1.11(3.11)	.19(.117)	0(0)
(p, q)	Method	Error($\widehat{C}$)	Error($\widehat{\Omega}$)	Error($\widehat{C}, \widehat{\Omega}$)	FPR($\widehat{\Omega}$)	FNR($\widehat{\Omega}$)
$(3, 200)$	Our-ML	3.01(.15)	3.58(.203)	6.77(.282)	.004(.001)	.004(.002)
	Our-ML-convex	3.01(.15)	8.99(.195)	11.9(.238)	.067(.006)	0(0)
	MRCE-group	3.02(.15)	37.1(4.8)	4.5(4.65)	.162(.08)	.48(.255)

References

J. Fan and R. Li. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*, 96(456):1348–1360, 2001.

X. Shen, W. Pan, and Y. Zhu. Likelihood-Based Selection and Sharp Parameter Estimation. *Journal of the American Statistical Association*, 107(497):223–232, June 2012.

C. H. Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942, 2010.