We cordially thank the reviewers for their time and thoughtful comments. We will improve the presentation of the paper and further develop the experiments, including testing on real data (which are abundant in the prior works on LASSO), as suggested. Moreover, some specific comments by the reviewers are addressed as follows:

Reviewer 1

- Regarding the appropriateness of our paper to neurips: We would like to mention that there have been previous papers in the same category which appeared on neurips (and were highly influential). For example, The Javanmard and Montanari paper we cited has a neurips 2013 version: "Confidence Intervals and Hypothesis Testing for High-Dimensional Statistical Models". We have plans of further developing the experiments, and submitting a longer version to journal for reference. Thank you for the suggestions.
- In Definition 1, indeed the noise scales with the number of samples: $N_i \sim \mathcal{N}(0, n\sigma^2)$. This scaling ensures that for each j, the square error in estimating θ_j scales as $\Theta(n)/n = \Theta(1)$ (where $1/n$ factor because of n samples), in which case we have hope of convergence of the empirical distribution of the error, as we desire.
- In Definition 1, indeed $\mathbf{X}_i \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma})$. We updated the manuscript.
- Example of constructing Gaussian knockoffs: in the Gaussian case it all boils down to designing the covariance matrix. eq 5-10 show the previous ways, and for conditional independence knockoff, the explict formula is given in (27).
- By "regressing Y on $[X^p, \tilde{X}^p]$", we mean solving $\min_{\theta=(\theta_1,\theta_2)}\{\|Y - [X^p, \tilde{X}^p]\theta\|^2 + \lambda\|\theta\|_1\}$, as opposed to solving $\min_{\theta_1}\{\|Y - X^p\theta_1\|^2 + \lambda\|\theta_1\|_1\}$ and $\min_{\theta_2}\{\|Y - \tilde{X}^p\theta_2\|^2 + \lambda\|\theta_2\|_1\}$ separately. We will further clarify this in the texts around (3).
- Assumption of $|\theta_j| \geq 1$: actually we can completely drop this assumption, using a more careful analysis, and the bounds (22) and (23) will be replaced by new bounds depending on the limit of the empirical distribution of $\theta_0^{(p)}$ (whose existence is guaranteed by the standard distributional limit assumption).
- Simply put, we found a simple necessary and sufficient condition on $\mathbf{\Sigma}$ for low FDR and high power: empirical distribution of $((\mathbf{\Sigma}^{-1})_{jj})_{j\in[p]}$ should converge to 0 in distribution. For example, since convergence in expectation implies convergence in distribution (Markov inequality), having $\frac{1}{p}\sum_{j=1}^{p}(\mathbf{\Sigma}^{-1})_{jj}$ bounded above is sufficient.

Reviewer 2

- Regarding the necessity of the condition for FDR$\to 0$ and POWER$\to 1$ for the knockoff filter: actually our condition $\|(\underline{P}_{jj})_{j=1}^{2p}\|_{LP} \to 0$ is not only sufficient but also necessary. The intuitive explanation is that $\hat{\theta}_j^u - \theta_{0,j}$ is roughly distributed as $\mathcal{N}(0, \tau^2\underline{P}_{jj}^{-1})$ (by the standard distributional limit), so that FDR$\to 0$ and POWER$\to 1$ if and only if the fraction of j's for which $\underline{P}_{jj}$ exceeds any given threshold asymptotically vanish (i.e., weak convergence of the empirical distribution of $(\underline{P}_{jj})_{j\in[2p]}$). The proof of converse will be similar to achievability. However, the reviewer is right that the we should make this point clearer in the revised version.
- Support recovery: the variable selection problem might be interpreted as support recovery. However, to our knowledge, literature on support recovery usually focuses on exactly recovering the whole support (e.g. Knight and Fu, "Asymptotics for lasso-type estimators," 2000, and Zhao and Yu, "On model selection consistency of Lasso," 2006). In contrast, FDR may be considered as a softer criterion for the quality of support recovery. Exact support recovery is not asymptotically feasible in the regime we consider. We will add discussions and related citations.
- Comparison with SDP knockoff: testing binary tree with various correlations yield similar simulation results, namely that the conditional independence knockoff performs similarly but slightly better than sdp knockoff. We plan to test other trees or sparse graphs. Notwithstanding, the conditional independence knockoff is still much more computationally efficient to construct than sdp knockoff, and more reliable since ESD has a closed-form expression.

Reviewer 3

- Regarding relevance to real data: assumptions such as sparse precision matrix or tree graphical models are very common, and in fact, many algorithms (such as Chow-Liu or Graphical Lasso) rely on such assumptions for any hope of estimating of the precision matrix. Also as noted above, we can drop condition (16) in Prop 4 and use the limiting empirical distribution of θ_0 instead. We will further test the algorithm in certain real data sets. A starting point might be similar data sets in the previous Lasso literature, such as those in the paper of Javanmard and Montanari. For sparse precision matrix, we are considering similar data as those found in the papers of Bühlmann, Kalisch, and Meier.