Cluster recovery under smooth field modulation with guarantees

1. **Summary and Contributions: Briefly summarize the paper and its contributions**

   This work tackles the disentanglement problem to recover the cluster assignment and level from composite data with additive smooth field signal. The work is motivated by the real MRI analysis in medical imagining. Under the noiseless setting, identifying conditions to exact recover the cluster assignment are characterized by the data connectivity, and deviation bound of cluster level estimation is provided. An alternating algorithm is also provided to solve the clustering problem.

2. **Strengths: Please describe the strengths of the work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   The work is well-motivated by the MRI bias field removal problem. The real data analysis of MRI tissue recovery is interesting.

3. **Weaknesses: Please describe the limitations of this work according (but not limited) to the following criteria: soundness of the claims (theoretical grounding, empirical evaluation), significance and novelty of the contribution, and relevance to the AISTATS community.**

   There is a gap between the provided theoretical guarantees and the practical applications. Specifically, all theoretical results are done under the ideal noiseless case, whereas, the guarantees for both statistical estimator (3) and algorithm (ALTMIN) are not provided under the noisy cases. Further, one may not see the outstanding empirical performance of the proposed method due the lack of numerical comparison with competitive methods. Therefore, it is hard for me to recognize the significance of proposed method in practice.

   See additional comments for other comments.

4. **Correctness: Are the method and claims correct? Is the empirical methodology correct?**

   Most parts of the paper seem sound for me. Some details in numerical experiments are unclear; see additional comments for details.

5. **Clarity: Is the paper well written? Does it clearly state its contributions, notation and results?**

   I feel that some parts of the writing could be improved with a more concise and clearer expression.

   For example, in page 1, the sentence "Of particular interest to us, will be the be the recovery ... signals" lacks subject and has a typo; the sentence "Note that the original data process

need not ..., only that, after an appropriate transformation, ... terms of (1)" may be organized in a clearer way.

6. **Relation to prior work: Is it clearly discussed how this work differs from or relates to prior work in the literature?**

   Numerical comparisons with competitive bias field removal problem are missing. Without the previous methods serving as benchmarks, it is hard to evaluate the performance of proposed algorithm compared with the state-of-art.

7. **Additional Comments: Add your additional comments, feedback and suggestions for improvement, as well as any further questions for the authors.**

   - The general selection criteria of the parameters M, λ and the kernel K are not provided. Adding more discussions for practical hyperparameter selection may be helpful.

   - Figure 3 seems inconsistent with the setting (11). When $M = 3$, we have $\mu_k^* \in \{1, 2, 3\}$ while Figure 3 has level $\mu_k^* \in \{0, 0.5, 1\}$.

   - The noisy cases consider the setting with very small variance $\sigma^2 \in \{0.01, 0.04, 0.09\}$. By Figure 5, the clustering performance drops quickly after adding slight perturbations. More discussions on the algorithm robustness against large noise and the signal-to-noise (maybe $\sigma^2 = 0.09$ is a large noise under the proposed model) may be helpful.

   - More details of the real data analysis should be provided, including the data sample size and hyperparameter selections.

References