

Response to Reviewer 2

January 20, 2017

We appreciate the helpful feedback from the reviewer. We have addressed your questions and comments. Below we give a point-by-point response to each of the questions posed by the reviewer:

1. Last paragraph of page 4: you assume that (3) for the training set is strictly convex in θ . I am wondering if the strict convexity assumption would exclude some interesting high-dimensional cases. Can this assumption be removed? Otherwise, it should be listed as a separate condition along with Conditions 1-3 on page 7, as it is a strong condition.

The reviewer is correct in noting that this assumption can be removed. The paper only requires that the training criterion is convex and has a unique minimizer. We have removed this condition from the paper.

2. Page 7, I found the statement of Definition 3 somewhat disconnected with the rest of the paper. For example, you assume the $n \times p$ matrix B has orthonormal columns. Does this require p is smaller than (or equal to) n ? Can you modify this definition directly using the subspace defined earlier?

The reviewer is correct in noting that Definition 3 uses misleading notation. n is supposed to be arbitrary number, not the training size. We have replaced n with a different variable q to clarify this.

3. Examples in Section 2.4. It was said that details are included in the Appendix. I found those details in the Appendix are still very brief. Could you provide more details on how to check Conditions 1-3 for each example in the Appendix? And how about the strong convexity assumption? Is it satisfied, too?

We have included more detail in the Appendix on how to check Conditions 1-3. Also, as noted in the response to the first question from the reviewer, we have removed the strong convexity assumption.

4. Examples in Section 3. Could you add false positive and false negative in each of the tables? And how about the computational speed? The p in each of the examples is still small. Could you provide some general comments on how well the new algorithm can handle large p ?

Section 4 was a classification problem, we have provided false positive and false negative rates in Table 6. For the examples in Section 3, it's unclear what false positive and false negative would mean since they are regression problems. In the previous submission, we included the percent of correctly identified nonzero features. We now realize that this metric is misleading, since minimizing the validation error via cross-validation is meant for finding penalty parameters that have small generalization error. It has been shown (CITE?) that selecting penalty parameters via cross-validation results in models that tend to have too many nonzero elements. That is, the penalty parameters give the correct active set (the set of nonzero elements) should not be chosen by cross-validation. Therefore we have removed this metric entirely from the paper.