Response to Reviewer Comments

Mario Coutino, Sundeep P. Chepuri, and Geert Leus

May 18, 2018

First of all, we would like to thank the reviewers for their thorough and critical review of our work, and in particular for providing useful suggestions aimed to increase the quality of the work and improve its clarity. In addition, we would like to thank the AE for handling our manuscript in such a timely and effective manner. Major changes in the manuscript have been highlighted. Following the suggestions of the reviewers and the comments of the AE we have made the following changes to the manuscript:

- the derivation in Theorem 2 has been clarified and the following minor addition has been made
  - the suggestion for the derivation of Reviewer 1 has been added
- Eq. $(53-55)$ have been modified to use a general set $\mathcal{B}$ instead of sets differing in a single element as recommended by Reviewer 1.
- the claims with respect to the complexity of the greedy algorithm have been made more precise

Comments from Reviewer 1

I once again thank the authors and editors for taking the time to address my concerns. There are minor points still outstanding, but the major issue is that there are still typos in the proof. Though I don't believe they would change the result much, I cannot recommend this paper for publication unless the proof of Theorem 2, the main theoretical result of the paper, is correct.

Response: We want to thank you for your comments in the previous rounds. They provided helped us improve the quality of the manuscript. We really appreciate the thorough proof reading that the reviewer performed over the different versions of our manuscript, particularly with respect to Theorem 2, as it has improved the quality, clarity and correctness of the paper.

Comment 1*: I refer here to the sequence of inequalities after (63). I believe the min in the third inequality does not hold. I believe the right sequence they are looking for is $\lambda_{min}^{-1}(a\boldsymbol{I} + \boldsymbol{\Phi S \Phi}) = [a + \lambda_{min}(\boldsymbol{\Phi S \Phi})]^{-1} <= [a + \lambda_{min}(S)]^{-1}$. If the authors use a different bound or approximation, it should be made clear in the text. Currently, the result reads $\lambda_{min}^{-1}(a\boldsymbol{I} + \boldsymbol{\Phi S \Phi}) <= \min(1/a, [(1-\beta)\lambda_{min}(\boldsymbol{\Sigma})]^{-1}) = \min(1/a, [\lambda_{min}(\boldsymbol{S})]^{-1})$, since $a = \beta\lambda_{min}(\boldsymbol{\Sigma})$ so that $(1-\beta)\lambda_{min}(\boldsymbol{\Sigma}) = \lambda_{min}(\boldsymbol{\Sigma}) - a = \lambda_{min}(\boldsymbol{S})$. I'd ask that the authors recheck their proof to make sure it holds and if so, give more details on these steps.*

Another thing is that (54) does not match the definition of epsilon-approximate submodularity given in Definition 2, since it involves only sets whose cardinalities differ by at most 1. I think that changing (53)-(55) to a generic $\mathcal{B}$ as in Definition 2 would not alter the result, but it would make the proof correct.

Response: We thank again the reviewer for its concerns with respect to the quality of our results. After your comment, we include here the rationale behind the inequalities in the manuscript:

$$\begin{aligned}
\lambda_{\min}^{-1}\{a\boldsymbol{I} + \boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\} &\leq \min\{\lambda_{\min}^{-1}\{a\boldsymbol{I}\}, \lambda_{\min}^{-1}\{\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\}\} && (1)\\
&\leq \min\{1/a, \lambda_{\min}^{-1}\{\boldsymbol{S}\}\} && (2)\\
&= \min\{1/a, (1-\beta)^{-1}\lambda_{\min}^{-1}\{\boldsymbol{\Sigma}\}\}, && (3)
\end{aligned}$$

where (1) follows from the composition of PD matrices, (2) uses the interlacing theorem for submatrices of PD matrices, and (3) restates the result in terms of the parameters a and β.

As the reviewer pointed out, an alternative route to obtain the inequalities is through its suggestion. That is, we can follow the chain

$$\begin{aligned} \lambda_{\min}^{-1}\{a\boldsymbol{I}+\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\} &\leq \lambda_{\min}^{-1}\{a\boldsymbol{I}+\boldsymbol{S}\} &(4)\\ &= \lambda_{\min}^{-1}\{\boldsymbol{\Sigma}\}, &(5)\end{aligned}$$

which in turns leads to a similar bound. Here, we notice that the bound using (5) might be tighter than using (3). However, as pointed out by the reviewer in previous rounds, using (5) does not include the parameters a, and β as explicitly as using (3). As a result, during the previous revision of the manuscript we decided to use the bound obtained with (3). Nonetheless, as the bound resulting from using (5) results in a *tighter* bound, we have decided to update our manuscript with the reviewer suggestion. This change does not affect our result. Furthermore, observing the corollary, when $\beta \to 0$, we notice that the bound in (67) does not change when (5) is used.

> [...] For the upper bound, we notice that by the maximum singular value of the second matrix, the following inequality holds
>
> $$\begin{aligned}\boldsymbol{\Delta} &\leq a\boldsymbol{I}+\sigma_{\max}\Big\{\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\Big(a\boldsymbol{I}+\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\Big)^{-1}\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\Big\}\boldsymbol{I}\\ &\leq a\boldsymbol{I}+\lambda_{\min}^{-1}\Big\{a\boldsymbol{I}+\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\boldsymbol{\Phi}_{\mathcal{A}}^T\Big\}\sigma_{\max}^2\Big\{\boldsymbol{\Phi}_{\mathcal{A}}\boldsymbol{S}\Big\}\boldsymbol{I}\\ &\leq a\boldsymbol{I}+\lambda_{\min}^{-1}\{a\boldsymbol{I}+\boldsymbol{S}\}\lambda_{\max}^2\{\boldsymbol{S}\}\boldsymbol{I},\\ &\leq (a+\nu\lambda_{\max}^2\{\boldsymbol{S}\})\boldsymbol{I},\end{aligned}$$
>
> where the submultiplicativity and subadditivity of singular values, and the interlacing theorem for submatrices of PD matrices are used in the second and third inequality, respectively, and we have defined $\nu = \lambda_{\min}^{-1}\{\boldsymbol{\Sigma}\}$. [...]

With respect to the reviewer comment of the expressions in equations (53–55) we have modified the equations (53 – 55) to match the expressions used in Def. 1 and Def. 2 that makes use of a general set $\mathcal{B}$. Here are the changes made in the manuscript:

> [...] Using (52) and considering $\mathcal{A} \subseteq \mathcal{B} \subseteq \mathcal{V}$ and $i \notin \mathcal{B}$, we can obtain the following expression
>
> $$\delta(\mathcal{A}\cup\{i\})-\delta(\mathcal{A})-\delta(\mathcal{B}\cup\{i\})+\delta(\mathcal{B}) \geq -4\epsilon', \tag{53}$$
>
> where we have defined $\delta(\mathcal{A}) = \tilde{s}(\mathcal{A}) - \hat{s}(\mathcal{A})$. Due to the modularity of $\hat{s}(\mathcal{A})$, i.e.,
>
> $$\hat{s}(\mathcal{A}\cup\{i\})-\hat{s}(\mathcal{A})-\hat{s}(\mathcal{B}\cup\{i\})+\hat{s}(\mathcal{B}) = 0, \tag{54}$$
>
> and using expressions (53) and (54), we can show that
>
> $$\tilde{s}(\mathcal{A}\cup\{i\})-\tilde{s}(\mathcal{A})-\tilde{s}(\mathcal{B}\cup\{i\})+\tilde{s}(\mathcal{B}) \geq -4\epsilon'. \tag{55}$$
>
> [...]

Comment 2: *The authors have a point that for fixed M or K, greedy search has linear complexity in the non-fixed parameter. However, both parameters are inputs to the algorithm and should be considered as such. The complexity is therefore polynomial. Claiming greedy search to be of linear complexity is incorrect. Check, for instance, [Sipser, Introduction to the theory of computation]. In fact, there has been work on constructing a linear time greedy algorithm, e.g., [Mirzasoleiman et al. Lazier than lazy greedy], but it involves randomization. Naturally, these are all query complexity results.*

Response: Thanks for pointing out this situation. Complexity notation for more than one input variable is always problematic. Due to this, we made the statement that when one of these two parameters is fixed, the greedy search has linear complexity. For sensor selection applications, this is the common case, e.g., fixed number of available sensors. We, under no circumstances, claim that it is a *linear complexity algorithm* for M and K as this is clearly not true. Therefore, we consider that stating that: *"greedy search scales linearly on the available set for fixed K"* or *"greedy search scales linearly on the number of selected sensors for fixed M"* is not incorrect as long as we are precise in the statement (in these cases, it is always assumed that the other parameter is fixed). However, in order to consider the reviewer comment we made our claims even more precise by using the following statements in the manuscript:

> [...] it accepts a near-optimal maximization using a greedy algorithm that despite its general polynomial complexity, scales linearly in the number of available sensors through its recursive description when the number of selected sensors is fixed [...]

> [...] are more attractive as for a fixed K the number of function evaluations required by the method scales linearly in the number of available sensors. Therefore, the complexity only depends on the efficient evaluation of the cost function. [...]

> [...] the proposed method, for a fixed K, scales linearly with respect to the number of available sensors. [...]

We understand your concern from the theoretical complexity perspective. Depending on the size of the input, notations such as $O(MK)$ might imply different things, e.g., matrix multiplication is considered *linear* as the input is of size $O(MK)$, while for string lookups $O(MK)$ implies $O(\max\{M, K\}^2)$ (the inputs are two strings which possible maximum length). For the case of Mirzasoleiman et al., it can be argued that the query complexity $O(M \log 1/\epsilon)$, in worst case, is also polynomial (with a very low exponent) and not linear. However, it is commonly accepted that ϵ is not related to the input and treated similar to a fixedparameters in fixed-parameter tractable (FPT) problems. In certain sense, the sampling strategy introduced in Mirzasoleiman et al creates another dependency ($\log 1/\epsilon$), which might be artificially related with K or M, but that is expected to be smaller than K for practical cases. This leads to the main motivation of Mirzasoleiman et al. which is the reduction of the dependency in K. This behaviour is obtained by degrading the near-optimality guarantees of the algorithm.