Appendix

A. Organization of the Appendix

Appendix B presents the proofs for our approximation guarantees and its tightness for the GREEDY algorithm.

Appendix C provides details on existing notions of curvature and submodularity ratio, and relates it to the notions in this paper.

Appendix D presents detailed proofs for bounding the submodularity ratio and curvature for various applications.

Appendix E gives details on the classical SDP formulation of the Bayesian A-optimality objective.

Appendix F provides proofs omitted in Section 6.

Appendix G provides information on more applications, including sparse modeling with strongly convex loss functions, subset selection using the R^2 objective and optimal budget allocation with combinatorial constraints.

Appendix H provides experimental results on subset selection with the R^2 objective and additional results on experimental design.

B. Proofs for Approximation Guarantee and Tightness Result (Section 2 and Section 3)

B.1. Proof of Remarks in Section 2

Proofs of Remark 1.

a) Because F is nondecreasing, and γ, γ^G are defined as the largest scalars, $\gamma, \gamma^G \geq 0$. At the same time, both γ and γ^G can be at most 1 because the conditions in Def. 1 also have to hold for the case that $\Omega \backslash S$ ($\Omega \backslash S^t$, respectively) is a singleton.

b) " $\Rightarrow$ ":

Let $\Omega \setminus S = \{\omega_1, \ldots, \omega_k\}, k \geq 1$. Submodularity implies $\sum_{i=1}^k \rho_{\omega_i}(S) \geq \rho_\Omega(S)$. Hence, γ can take the largest value 1.

" $\Leftarrow$ ":

$\gamma = 1$ implies that (setting $\Omega \setminus S = \{\omega_i, \omega_j\}$), for all $\omega_i, \omega_j \in \mathcal{V} \setminus S$, it holds that $F(\{\omega_i\} \cup S) + F(\{\omega_j\} \cup S) \geq F(\{\omega_i, \omega_j\} \cup S) + F(S)$, which is an equivalent way to define submodularity (Bach, 2013, Proposition 2.3).

□

Proof of Remark 2.

a) "If $F(\cdot)$ is nondecreasing, then $\alpha, \alpha^G \in [0, 1]$";

When $\Omega = \emptyset$, α is at least 0. From the definition, $\alpha^G \geq 0$. Since F is nondecreasing, $\rho_i(S \setminus \{i\} \cup \Omega) \geq 0$ (respectively, $\rho_{j_i}(S^{i-1} \cup \Omega) \geq 0$), and we defined α, α^G to be the smallest scalar, it must hold that $\alpha, \alpha^G \leq 1$.

b) "For a nondecreasing function $F(\cdot)$, $F(\cdot)$ is supermodular iff $\alpha = 0$ ";

" $\Rightarrow$ ":

If F is supermodular, it always holds that $\rho_i(S \setminus \{i\} \cup \Omega) \geq \rho_i(S \setminus \{i\})$, combined with the fact that α is at least 0, we know that α must be 0.

" $\Leftarrow$ ":

One can observe that $\alpha = 0$ is equivalent to $-F(\cdot)$ satisfying the diminishing returns property, which is equivalent to $F(\cdot)$ being supermodular.

c) "If $F(\cdot)$ is nondecreasing submodular, then $\alpha^G \leq \alpha = \alpha^{\text{total}}$."

Since it always holds that $\alpha^G \leq \alpha$, we only need to prove that $\alpha = \alpha^{\text{total}}$. Wlog., assume $\rho_i(S \setminus \{i\}) > 0$. Then,

$$\begin{aligned}
1 - \alpha &= \min_{\Omega, S \subseteq \mathcal{V}, i \in S \setminus \Omega} \frac{\rho_i(S \setminus \{i\} \cup \Omega)}{\rho_i(S \setminus \{i\})} \\
&= \min_{S \subseteq \mathcal{V}, i \in S} \frac{\rho_i(\mathcal{V} \setminus \{i\})}{\rho_i(S \setminus \{i\})} \quad \text{(diminishing returns, and taking } \Omega = \mathcal{V} \setminus \{i\}) \\
&= \min_{i \in \mathcal{V}} \frac{\rho_i(\mathcal{V} \setminus \{i\})}{\rho_i(\emptyset)} \quad \text{(diminishing returns, and taking } S = \{i\}) \\
&= 1 - \alpha^{\text{total}}.
\end{aligned}$$

So it holds that $\alpha^G \leq \alpha = \alpha^{\text{total}}$. □

B.2. Proof of Lemma 1

Proof of Lemma 1. The proof needs the definitions of generalized curvature, submodularity ratio, and the selection rule of the GREEDY algorithm.

Firstly, observe,

$$\begin{aligned}
F(\Omega \cup S^t) &= F(\Omega) + \sum_{i: j_i \in S^t} \rho_{j_i}(\Omega \cup S^{i-1}) \\
&= F(\Omega) + \sum_{i: j_i \in S^t \setminus \Omega} \rho_{j_i}(\Omega \cup S^{i-1}) + \underbrace{\sum_{i: j_i \in S^t \cap \Omega} \rho_{j_i}(\Omega \cup S^{i-1})}_{= 0 \text{ because } j_i \in \Omega} \\
&= F(\Omega) + \sum_{i: j_i \in S^t \setminus \Omega} \rho_{j_i}(\Omega \cup S^{i-1}). \qquad (7)
\end{aligned}$$

From the definition of the submodularity ratio,

$$F(\Omega \cup S^t) \leq F(S^t) + \frac{1}{\gamma} \sum_{\omega \in \Omega \setminus S^t} \rho_\omega(S^t). \qquad (8)$$

From the definition of curvature (for the greedy curvature, since it holds for S^{K-1}, it must also hold for $S^t \subseteq S^{K-1}$), we have,

$$\sum_{i: j_i \in S^t \setminus \Omega} \rho_{j_i}(\Omega \cup S^{i-1}) \geq (1 - \alpha) \sum_{i: j_i \in S^t \setminus \Omega} \rho_{j_i}(S^{i-1}). \qquad (9)$$

Combining (7) to (9), and remember that we use the shorthand $\rho_t := \rho_{j_t}(S^{t-1})$, it reads,

$$\begin{aligned}
F(\Omega) &= F(\Omega \cup S^t) - \sum_{i: j_i \in S^t \setminus \Omega} \rho_{j_i}(\Omega \cup S^{i-1}) \\
&\leq \alpha \sum_{i: j_i \in S^t \setminus \Omega} \rho_i + F(S^t) - \sum_{i: j_i \in S^t \setminus \Omega} \rho_i + \frac{1}{\gamma} \sum_{\omega \in \Omega \setminus S^t} \rho_\omega(S^t) \\
&= \alpha \sum_{i: j_i \in S^t \setminus \Omega} \rho_i + \sum_{i: j_i \in S^t \cap \Omega} \rho_i + \frac{1}{\gamma} \sum_{\omega \in \Omega \setminus S^t} \rho_\omega(S^t) \\
&\leq \alpha \sum_{i: j_i \in S^t \setminus \Omega} \rho_i + \sum_{i: j_i \in S^t \cap \Omega} \rho_i + \gamma^{-1}(K - w^t)\rho_{t+1},
\end{aligned}$$

where the last inequality is because of the selection rule of the GREEDY algorithm ($\rho_\omega(S^t) \leq \rho_{t+1}, \forall \omega$). □

B.3. Proof of Claim 1

Since the proof heavily relies on the structure of the constructed LPs, we restate it here: The worst-case approximation ratio of the group $\mathcal{P}_{K,\alpha,\gamma}(\{l_1,...,l_s\})$ is

$$\underline{R}(\{l_1,...,l_s\}) = \min \sum_{i=1}^{K} x_i, \text{ s.t. } x_i \geq 0, i \in [K] \text{ and}$$

$$\begin{matrix} \text{row }(0) \\ \text{row }(1) \\ \vdots \\ \text{row }(l_1-1) \\ \text{row }(l_2-1) \\ \text{row }(q=l_r) \\ \vdots \\ \text{row }(l_s-1) \\ \vdots \\ \text{row }(K-1) \end{matrix} \begin{bmatrix} K/\gamma & & & & & & & & & \\ \alpha & K/\gamma & & & & & & & & \\ \vdots & \vdots & \ddots & & & & & & & \\ \alpha & \alpha & \cdots & K/\gamma & & & & \mathbf{0} & & \\ \alpha & \alpha & \cdots & 1 & (K-1)/\gamma & & & & & \\ \alpha & \alpha & \cdots & 1 & 1 & \frac{K-r}{\gamma} & & & & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & & & \\ \alpha & \alpha & \cdots & 1 & 1 & \alpha & \cdots & \frac{K-s+1}{\gamma} & & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \ddots & \\ \alpha & \alpha & \cdots & 1 & 1 & \alpha & \cdots & 1 & \cdots & \frac{K-s}{\gamma} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_{l_1} \\ x_{l_2} \\ x_{q+1} \\ \vdots \\ x_{l_s} \\ \vdots \\ x_K \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \quad (2)$$

For notational simplicity, w.l.o.g., assume that $j_i = i, i \in [K]$. Let the row index in (2) start from 0.

Proof of Claim 1.

Let $\epsilon = \frac{(K-r+1-\gamma)x_q^* - (K-r)x_{q+1}^*}{K-r+1} \geq \frac{K-r}{K-r+1}(x_q^* - x_{q+1}^*) > 0$.

a) It is easy to see that $\boldsymbol{y}^* \geq 0$ since the only decreased entry is the q^{th} entry, and one can easily see that $y_q^* = x_q^* - \epsilon \geq 0$.

b) "All of the constraints in (2) are still feasible for $\boldsymbol{y}^*$."

(i) For the rows 0 to $(q-2)$ in (2), there is no change, so they are still feasible.

(ii) For the $(q-1)^{\text{th}}$ and q^{th} rows in (2), they are

$$\alpha x_1^* + \cdots + \alpha(\text{or } 1)x_{q-1}^* + \frac{K-r+1}{\gamma}x_q^* \geq 1 \tag{10}$$

$$\alpha x_1^* + \cdots + \alpha(\text{or } 1)x_{q-1}^* + x_q^* + \frac{K-r}{\gamma}x_{q+1}^* \geq 1 \tag{11}$$

For (10), after plugging $\boldsymbol{y}^*$ into its L.H.S., we get $\alpha y_1^* + \cdots + \alpha(\text{or } 1)y_{q-1}^* + \frac{K-r+1}{\gamma}y_q^*$, subtract from which the L.H.S. of (11), we get

$$\begin{aligned} &\left[\alpha y_1^* + \cdots + \alpha(\text{or } 1)y_{q-1}^* + \frac{K-r+1}{\gamma}y_q^*\right] - \left[\alpha x_1^* + \cdots + \alpha(\text{or } 1)x_{q-1}^* + x_q^* + \frac{K-r}{\gamma}x_{q+1}^*\right] \\ &= \frac{K-r+1}{\gamma}(x_q^* - \epsilon) - x_q^* - \frac{K-r}{\gamma}x_{q+1}^* \\ &= 0, \end{aligned}$$

so $\alpha y_1^* + \cdots + \alpha(\text{or } 1)y_{q-1}^* + \frac{K-r+1}{\gamma}y_q^* \geq 1$ and $\boldsymbol{y}^*$ is feasible for (10).

After increasing x_{q+1}^* by $\epsilon_{q+1} = \epsilon\frac{\gamma}{K-r}$, the q^{th} row in (2) is feasible since the change in its L.H.S. is $-\epsilon + \epsilon = 0$.

(iii) For the rows q to $(K-1)$ in (2), let us prove by **induction**.

For the base case, consider the $(q+1)^{\text{th}}$ row in (2), it can be either,

$$\alpha x_1^* + \cdots + \alpha(\text{or } 1)x_{q-1}^* + x_q^* + x_{q+1}^* + \frac{K-r-1}{\gamma}x_{q+2}^* \geq 1$$

or

$$\alpha x_1^* + \cdots + \alpha (\text{or } 1) x_{q-1}^* + x_q^* + \alpha x_{q+1}^* + \frac{K-r}{\gamma} x_{q+2}^* \geq 1$$

It can be easily verified that the $(q+1)^{\text{th}}$ row in (2) is still feasible in both the above two situations. Let us use Δ_{q+u} to denote the change of L.H.S. of the $(q+u)^{\text{th}}$ row after applying the changes.

For the inductive step, assume that the claim holds for $u = u'$, i.e., the $(q+u')^{\text{th}}$ row in (2) is feasible or $\Delta_{q+u'} \geq 0$. The $(q+u')^{\text{th}}$ row is,

$$(\texttt{...same as } (q+u'+1)^{\texttt{th}} \texttt{ row}) \; + \frac{K-r-v}{\gamma} x_{q+u'+1}^* \geq 1$$

where $0 \leq v \leq u'$ is some integer dependent on the structure of (2), but not affect the final analysis. Then the $(q+u'+1)^{\text{th}}$ row can be either,

$$(\texttt{... same as } (q+u')^{\texttt{th}} \texttt{ row}) \; + x_{q+u'+1}^* + \frac{K-r-v-1}{\gamma} x_{q+u'+2}^* \geq 1 \; \texttt{(case 1)}$$

or

$$(\texttt{... same as } (q+u')^{\texttt{th}} \texttt{ row}) \; + \alpha x_{q+u'+1}^* + \frac{K-r-v}{\gamma} x_{q+u'+2}^* \geq 1 \; \texttt{(case 2)}$$

In `(case 1)`, the L.H.S. of $(q+u'+1)^{\text{th}}$ row minus the L.H.S. of $(q+u')^{\text{th}}$ row is $\frac{K-r-v-1}{\gamma} x_{q+u'+2}^* - \frac{K-r-v-\gamma}{\gamma} x_{q+u'+1}^*$, so

$$\begin{aligned}
\Delta_{q+u'+1} - \Delta_{q+u'} &= \frac{K-r-v-1}{\gamma} \epsilon_{q+u'+2} - \frac{K-r-v-\gamma}{\gamma} \epsilon_{q+u'+1} \\
&= \left[\frac{K-r-v-1}{\gamma} \frac{K-r-u'-\gamma}{K-r-u'-1} - \frac{K-r-v-\gamma}{\gamma} \right] \epsilon_{q+u'+1} \\
&= \left[(K-r-v-1) \frac{K-r-u'-\gamma}{K-r-u'-1} - (K-r-v-\gamma) \right] \frac{\epsilon_{q+u'+1}}{\gamma} \\
&\geq \left[(K-r-v-1) \frac{K-r-v-\gamma}{K-r-v-1} - (K-r-v-\gamma) \right] \frac{\epsilon_{q+u'+1}}{\gamma} \qquad \texttt{(since } 0 \leq v \leq u'\texttt{)} \\
&= 0.
\end{aligned}$$

so the $(q+u'+1)^{\text{th}}$ row is still feasible.

In `(case 2)`, the L.H.S. of $(q+u'+1)^{\text{th}}$ row minus the L.H.S. of $(q+u')^{\text{th}}$ row is $\frac{K-r-v}{\gamma} x_{q+u'+2}^* - (\frac{K-r-v}{\gamma} - \alpha) x_{q+u'+1}^*$, so

$$\begin{aligned}
\Delta_{q+u'+1} - \Delta_{q+u'} &= \frac{K-r-v}{\gamma} \epsilon_{q+u'+2} - (\frac{K-r-v}{\gamma} - \alpha) \epsilon_{q+u'+1} \\
&\geq \frac{K-r-v}{\gamma} (\epsilon_{q+u'+2} - \epsilon_{q+u'+1}) \quad \texttt{(since } \alpha \geq 0\texttt{)} \\
&\geq 0. \quad \texttt{(since } \epsilon_{q+u'+2} \geq \epsilon_{q+u'+1}\texttt{)}
\end{aligned}$$

so the $(q+u'+1)^{\text{th}}$ row is feasible. Thus we finish proving Claim 1. □

B.4. Proof of Claim 2

Proof of Claim 2. The change of the LP objective is

$$\begin{aligned}
\Delta_{LP} &= -\epsilon + \epsilon_{q+1} + \epsilon_{q+2} + \cdots + \epsilon_K \\
&= \epsilon \left[-1 + \frac{\gamma}{K-r} + \frac{\gamma}{K-r} \cdot \frac{K-r-\gamma}{K-r-1} + \cdots + \frac{\gamma}{K-r} \cdot \frac{K-r-\gamma}{K-r-1} \cdots \frac{K-r-m+2-\gamma}{K-r-m+1} \right],
\end{aligned}$$

where inside the bracket there are $m = K - q$ items except for the "-1". For notational simplicity, let the sum inside the bracket to be,

$$h_r(\gamma) := -1 + \frac{\gamma}{K-r} + \frac{\gamma}{K-r}\cdot\frac{K-r-\gamma}{K-r-1} + \cdots + \frac{\gamma}{K-r}\cdot\frac{K-r-\gamma}{K-r-1}\cdots\frac{K-r-m+2-\gamma}{K-r-m+1}. \tag{12}$$

First of all, since $K - r \geq K - q = m$, we have that

$$h_r(\gamma) \leq h_{r=q}(\gamma) = -1 + \frac{\gamma}{m} + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1} + \cdots + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1}\cdots\frac{3-\gamma}{2}\cdot\frac{2-\gamma}{1}. \tag{13}$$

Let us merge the items in (13) from left to right one by one,

$$\begin{aligned} h_{r=q}(\gamma) &= -1 + \frac{\gamma}{m} + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1} + \cdots + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1}\cdots\frac{3-\gamma}{2}\cdot\frac{2-\gamma}{1} \\ &= -\frac{m-\gamma}{m} + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1} + \cdots + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1}\cdots\frac{3-\gamma}{2}\cdot\frac{2-\gamma}{1} \\ &= -\frac{m-\gamma}{m}\frac{m-1-\gamma}{m-1} + \cdots + \frac{\gamma}{m}\cdot\frac{m-\gamma}{m-1}\cdots\frac{3-\gamma}{2}\cdot\frac{2-\gamma}{1} \\ &\cdots \\ &= -\frac{(m-\gamma)(m-\gamma-1)\cdots(2-\gamma)(1-\gamma)}{m(m-1)\cdots 2\cdot 1} \\ &\overset{\text{setting } \gamma \text{ to be } 1}{\leq} 0 \end{aligned}$$

Then $h_r(\gamma) \leq 0, \forall \gamma \in (0, 1]$. And it is easy to see that the equality holds if $r = q$ and $\gamma = 1$.

So we have that $\Delta_{LP} = \epsilon h_r(\gamma) \leq 0$, where the equality is achieved at "boundary" situation ($r = q$ and $\gamma = 1$). □

B.5. Proof of Lemma 3

Proof of Lemma 3.

For notational simplicity, wlog., assume that $j_i = i, i \in [K]$.

a) Firstly let us prove that $\underline{R}(\{l_1, ..., l_s\}) \geq \underline{R}(\emptyset)$.

The high-level idea is to change the structure of the constraint matrix in the LP associated with $\{l_1, ..., l_s\}$, such that in each change, the optimal LP objective value $\underline{R}$ never increases.

To better explain the proof, let us state the *setup* first of all. Let us call the elements inside the set $\Omega^* \cap S^K = \{l_1 = j_{m_1}, l_2 = j_{m_2}, ..., l_s = j_{m_s}\}$ the "joint elements", which means that they are joint elements in Ω^* and S^K. Similarly, the elements outside of $\Omega^* \cap S^K$ are called the "disjoint" elements. For the joint elements, two elements l_i, l_j being "adjacent" means that $l_i + 1 = l_j$. Mapping to the constraint matrix in (2), it means that the corresponding columns (column (l_i) and column (l_j)) are adjacent with each other. So we also call the corresponding columns in the constraint matrix as "joint columns".

We prove part **a)** of Lemma 3 by two steps: In the first step, we try to make all of the joint elements inside $\{l_1, l_2, ..., l_s\}$ to be adjacent with each other; In the second step, we get rid of the joint columns in the constraint matrix from left to right, one by one. Specifically,

Step 1. Assume that some elements inside $\{l_1, l_2, ..., l_s\}$ are not adjacent, like the example in (2), where l_2 and l_3 are not adjacent. Suppose that l_r and l_{r+1} are not adjacent, which means $l_r + 1 < l_{r+1}$. Denote $p = l_r$ for notational simplicity. Let us use $\mathbf{A}$ to represent the constraint matrix in the constructed LP associated with $\{l_1, l_2, ..., l_{r-1}, l_r, l_{r+1}, ..., l_s\}$, let $\mathbf{A}'$ represent the constraint matrix associated with $\{l_1, l_2, ..., l_{r-1}, l_r + 1, l_{r+1}, ..., l_s\}$. Notice that $l_r + 1$ is a disjoint element for $\mathbf{A}$, but a joint element for $\mathbf{A}'$. Furthermore $\mathbf{A}$ and $\mathbf{A}'$ only differ by columns p and $p + 1 = l_r + 1$. Assume that $\boldsymbol{x}^* \in \mathbb{R}_+^K$ is the optimal solution of the constructed LP with $\mathbf{A}$ as its constraint matrix. From Lemma 2, it must hold that $x_p^* \leq x_{p+1}^*$. Combining with the fact that $\mathbf{A}\boldsymbol{x}^* \geq 1$, one can easily verify that $\mathbf{A}'\boldsymbol{x}^* \geq 1$, which implies that,

$$\underline{R}(\{l_1, l_2, ..., l_{r-1}, l_r, l_{r+1}, ..., l_s\})$$

$$\geq \underline{R}(\{l_1, l_2, ..., l_{r-1}, l_r + 1, l_{r+1}, ..., l_s\}). \tag{14}$$

The change from $\{l_1, l_2, ..., l_{r-1}, l_r, l_{r+1}, ..., l_s\}$ to $\{l_1, l_2, ..., l_{r-1}, l_r + 1, l_{r+1}, ..., l_s\}$ is essentially to swap the roles of one originally disjoint element $l_r + 1$ and the originally joint element l_r. Repeatedly applying this operation for all $1 \leq r \leq s-1$ such that $l_r + 1 < l_{r+1}$, we can get that,

$$\begin{aligned} &\underline{R}(\{l_1, l_2, ..., l_{r-1}, l_r, l_{r+1}, ..., l_s\}) \\ &\geq \underline{R}(\{l_s - s + 1, l_s - s + 2, ..., l_s - 1, l_s\}). \end{aligned} \tag{15}$$

Now the s joint elements inside $\{l_s - s + 1, l_s - s + 2, ..., l_s - 1, l_s\}$ are adjacent with each other.

Step 2. Let $\mathbf{B}$ be the constraint matrix associated with $\{l_s - s + 1, l_s - s + 2, ..., l_s - 1, l_s\}$, and $\mathbf{B}'$ be the constraint matrix associated with $\{l_s - s + 2, ..., l_s - 1, l_s\}$. Note that $\mathbf{B}$ and $\mathbf{B}'$ differ in the columns from $l_s - s + 1$ to the end. Suppose the vector $\boldsymbol{x}^*$ is the optimal solution of the constructed LP with $\mathbf{B}$ as the constraint matrix. According to Lemma 2 we know that $x^*_{l_s-s+1} \leq x^*_{l_s-s+2} \leq \cdots \leq x^*_{l_s} \leq x^*_{l_s+1}$. So one can easily verify that it must hold that $\mathbf{B}'\boldsymbol{x}^* \geq 1$, which implies

$$\begin{aligned} &\underline{R}(\{l_s - s + 1, l_s - s + 2, ..., l_s - 1, l_s\}) \\ &\geq \underline{R}(\{l_s - s + 2, ..., l_s - 1, l_s\}). \end{aligned} \tag{16}$$

Apply this process repeatedly s times, one can reach that $\underline{R}(\{l_s - s + 1, l_s - s + 2, ..., l_s - 1, l_s\}) \geq \underline{R}(\emptyset)$.

Combining step 1 and step 2, we prove part **a)** of Lemma 3.

b) Then let us prove that $\underline{R}(\emptyset) = \frac{1}{\alpha}\left[1 - \left(\frac{K-\alpha\gamma}{K}\right)^K\right]$.

The constructed LP associated with $\underline{R}(\emptyset)$ is,

$$\underline{R}(\emptyset) = \min \sum_{i=1}^{K} x_i$$

subject to the constraints that,

$$x_i \geq 0, \forall i = 1, ..., K$$

and

$$\begin{bmatrix} \frac{K}{\gamma} & & & & & & & & & \\ \alpha & \frac{K}{\gamma} & & & & & & & & \\ \vdots & \vdots & \ddots & & & & & & & \\ \alpha & \alpha & \cdots & \frac{K}{\gamma} & & & & \mathbf{0} & & \\ \alpha & \alpha & \cdots & \alpha & \frac{K}{\gamma} & & & & & \\ \alpha & \alpha & \cdots & \alpha & \alpha & \frac{K}{\gamma} & & & & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & \ddots & & & \\ \alpha & \alpha & \cdots & \alpha & \alpha & \alpha & \cdots & \frac{K}{\gamma} & & \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & \ddots & \\ \alpha & \alpha & \cdots & \alpha & \alpha & \alpha & \cdots & \alpha & \cdots & \frac{K}{\gamma} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_a \\ x_b \\ x_c \\ \vdots \\ x_d \\ \vdots \\ x_K \end{bmatrix} \geq \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \tag{17}$$

One can observe that the vector $\boldsymbol{y} \in \mathbb{R}^K_+$ such that $y_i = \frac{\gamma}{K}\left(\frac{K-\gamma\alpha}{K}\right)^{i-1}, i = 1, ..., K$ satisfies all the constraints and every row in (17) is tight, hence $\boldsymbol{y}$ is the optimal solution. So

$$\underline{R}(\emptyset) = \sum_{i=1}^{K} y_i = \frac{1}{\alpha}\left[1 - \left(\frac{K - \alpha\gamma}{K}\right)^K\right].$$

□

B.6. Proof for the Tightness Result

Proof of Lemma 4.

a) "When $\alpha = 0$, $F(\cdot)$ is supermodular";

It is easy to see that $\xi_i = 1/K, i \in [K]$. Since $f(\cdot)$ is convex, it can be easily verified that $F(\cdot)$ is supermodular.

b) "When $\gamma = 1$, $F(\cdot)$ is submodular";

Now $f(x) = x$. Assume there are $T_1 \subseteq T_2 \subseteq \mathcal{V}, t \in \mathcal{V} \backslash T_2$. Let $T_1 = S_1' \cup \Omega_1', T_2 = S_2' \cup \Omega_2'$, where $S_1', S_2' \subseteq S, \Omega_1', \Omega_2' \subseteq \Omega$. It holds that $S_1' \subseteq S_2', \Omega_1' \subseteq \Omega_2'$. Now there are two cases:

1) $t = j_i \in S$. Then,

$$\rho_{j_i}(T_1) = \left[1 - \frac{\alpha\gamma}{K} f(|\Omega_1'|)\right] \xi_i, \qquad \rho_{j_i}(T_2) = \left[1 - \frac{\alpha\gamma}{K} f(|\Omega_2'|)\right] \xi_i$$

Because $f(\cdot)$ is nondecreasing, so it holds $\rho_{j_i}(T_1) \geq \rho_{j_i}(T_2)$.

2) $t = \omega_i \in \Omega$. It reads,

$$\rho_{\omega_i}(T_1) = \frac{1}{K}\left[1 - \alpha\gamma \sum_{j_i \in S_1'} \xi_i\right], \qquad \rho_{\omega_i}(T_2) = \frac{1}{K}\left[1 - \alpha\gamma \sum_{j_i \in S_2'} \xi_i\right]$$

Because $S_1' \subseteq S_2'$, so $\rho_{\omega_i}(T_1) \geq \rho_{\omega_i}(T_2)$.

The above two situations prove the submodularity of $F(T)$ when $\gamma = 1$.

c) "$F(T)$ has submodularity ratio γ and curvature α".

Let us assume $T = A \cup B$ and $T' = A' \cup B'$ are two *disjoint* sets ($T \cap T' = \emptyset$), where A and A' are subsets of S while B and B' are subsets of Ω. It is easy to see that $A \cap A' = \emptyset, B \cap B' = \emptyset$.

First of all, for the **submodularity ratio**, assume without loss of generality[8] that $\rho_{T'}(T) > 0$, so the submodularity ratio is $\gamma = \min_{T,T'} \frac{\sum_{i \in T'} \rho_i(T)}{\rho_{T'}(T)}$.

One can see that,

$$\begin{aligned}
\rho_{T'}(T) &= F(T' \cup T) - F(T) \\
&= \frac{f(|B \cup B'|) - f(|B|)}{K}(1 - \alpha\gamma \sum_{j_i \in A} \xi_i) + \left[1 - \frac{\alpha\gamma}{K} f(|B \cup B'|)\right] \sum_{j_i \in A'} \xi_i
\end{aligned}$$

and

$$\begin{aligned}
\sum_{i \in T'} \rho_i(T) &= \sum_{\omega_i \in B'} \rho_{\omega_i}(T) + \sum_{j_i \in A'} \rho_{j_i}(T) \\
&= |B'| \frac{f(|B| + 1) - f(|B|)}{K} \left(1 - \alpha\gamma \sum_{j_i \in A} \xi_i\right) + \left[1 - \frac{\alpha\gamma}{K} f(|B|)\right] \sum_{j_i \in A'} \xi_i.
\end{aligned}$$

Because $f(|B|) \leq f(|B \cup B'|)$, so one has $\left[1 - \frac{\alpha\gamma}{K} f(|B \cup B'|)\right] \sum_{j_i \in A'} \xi_i \leq \left[1 - \frac{\alpha\gamma}{K} f(|B|)\right] \sum_{j_i \in A'} \xi_i$, equality holds when $B' = \emptyset$ or $A' = \emptyset$. Therefore,

$$\begin{aligned}
\frac{\sum_{i \in T'} \rho_i(T)}{\rho_{T'}(T)} &= \frac{|B'| \frac{f(|B|+1) - f(|B|)}{K} \left(1 - \alpha\gamma \sum_{j_i \in A} \xi_i\right) + \left[1 - \frac{\alpha\gamma}{K} f(|B|)\right] \sum_{j_i \in A'} \xi_i}{\frac{f(|B \cup B'|) - f(|B|)}{K}(1 - \alpha\gamma \sum_{j_i \in A} \xi_i) + \left[1 - \frac{\alpha\gamma}{K} f(|B \cup B'|)\right] \sum_{j_i \in A'} \xi_i} \\
&\geq \frac{|B'| \frac{f(|B|+1) - f(|B|)}{K} \left(1 - \alpha\gamma \sum_{j_i \in A} \xi_i\right) + \left[1 - \frac{\alpha\gamma}{K} f(|B|)\right] \sum_{j_i \in A'} \xi_i}{\frac{f(|B \cup B'|) - f(|B|)}{K}(1 - \alpha\gamma \sum_{j_i \in A} \xi_i) + \left[1 - \frac{\alpha\gamma}{K} f(|B|)\right] \sum_{j_i \in A'} \xi_i}
\end{aligned}$$

[8] If $\rho_{T'}(T) = 0$, from monotonicity of $F(\cdot)$, it must hold $\sum_{i \in T'} \rho_i(T) = 0$, this case is not of interest in Def. 1.

$$\geq \frac{|B'|(f(|B|+1)-f(|B|))}{f(|B\cup B'|)-f(|B|)}, \tag{18}$$

where (18) comes from the fact: $f(\cdot)$ is convex and nondecreasing in $[0,K]$, thus $|B'|\frac{f(|B|+1)-f(|B|)}{K}\left(1-\alpha\gamma\sum_{j_i\in A}\xi_i\right)\leq \frac{f(|B\cup B'|)-f(|B|)}{K}(1-\alpha\gamma\sum_{j_i\in A}\xi_i)$.

Now to continue with (18), one can verify that by setting $B=\emptyset, B'=\Omega$, the minimum of (18) is achieved as γ, thus proving the submodularity ratio to be γ.

Then for the **curvature**, for any $t\in T=A\cup B$, we want to lower bound $\frac{\rho_t(T\setminus\{t\}\cup T')}{\rho_t(T\setminus\{t\})}$. There are two cases:

1) When $t=j_i\in A$, we have

$$\begin{aligned}\frac{\rho_{j_i}(T\setminus\{j_i\}\cup T')}{\rho_{j_i}(T\setminus\{j_i\})} &= \frac{\left[1-\frac{\alpha\gamma}{K}f(|B\cup B'|)\right]\xi_i}{\left[1-\frac{\alpha\gamma}{K}f(|B|)\right]\xi_i}\\ &= \frac{1-\frac{\alpha\gamma}{K}f(|B\cup B'|)}{1-\frac{\alpha\gamma}{K}f(|B|)}. \end{aligned}\tag{19}$$

Since $f(\cdot)$ is convex and nondecreasing in $[0,K]$, it is easy to see that the minimum of (19) is achieved when $B=\emptyset, B'=\Omega$ as $1-\alpha$.

2) When $t=\omega_i\in B$, we have,

$$\begin{aligned}\frac{\rho_{\omega_i}(T\setminus\{\omega_i\}\cup T')}{\rho_{\omega_i}(T\setminus\{\omega_i\})} &= \frac{\frac{f(|B\cup B'|)-f(|B\cup B'|-1)}{K}\left[1-\alpha\gamma\sum_{i'\in A\cup A'}\xi_{i'}\right]}{\frac{f(|B|)-f(|B|-1)}{K}\left[1-\alpha\gamma\sum_{i\in A}\xi_i\right]}\\ &\geq \frac{1-\alpha\gamma\sum_{i'\in A\cup A'}\xi_{i'}}{1-\alpha\gamma\sum_{i\in A}\xi_i} \qquad (20)\\ &= \frac{1-\alpha+\alpha-\alpha\gamma\sum_{i'\in A\cup A'}\xi_{i'}}{1-\alpha\gamma\sum_{i\in A}\xi_i} \qquad (21)\end{aligned}$$

where (20) is because $f(\cdot)$ is convex and nondecreasing in $[0,K]$.

Since $\alpha-\alpha\gamma\sum_{i'\in A\cup A'}\xi_{i'}\geq 0$ and $-\alpha\gamma\sum_{i\in A}\xi_i\leq 0$, continuing with (21) we have,

$$\frac{\rho_{\omega_i}(T'\setminus\{\omega_i\}\cup T)}{\rho_{\omega_i}(T\setminus\{\omega_i\})}\geq 1-\alpha.$$

The above two cases jointly prove that the objective in (3) has curvature α. □

C. Existing Notions of Curvature and Submodularity Ratio

In this section we firstly discuss existing notions of curvature and submodularity ratio, then secondly we present the relations to the notions in this paper.

C.1. Classical Notions of Curvature and Submodularity Ratio

The curvature of submodular functions measures how close a submodular set function is to being modular, and has been used to prove improved theoretical results for constrained submodular minimization and learning of submodular functions (Iyer et al., 2013). Earlier, it has been used to tighten bounds for submodular maximization subject to a cardinality constraint (Conforti & Cornuéjols, 1984) or a matroid constraint (Vondrák, 2010).

Definition 3 (Curvature of submodular functions (Conforti & Cornuéjols, 1984; Vondrák, 2010; Iyer et al., 2013))**.** *The* total curvature κ_F *(which we term as* α^{total} *in the main text) of a submodular function* F *and the* curvature $\kappa_F(S)$ *w.r.t. a set* $S\subseteq\mathcal{V}$ *are defined as,*

$$\kappa_F := 1-\min_{j\in\mathcal{V}}\frac{\rho_j(\mathcal{V}\setminus\{j\})}{\rho_j(\emptyset)} \text{ and}$$

$$\kappa_F(S) := 1 - \min_{j\in S} \frac{\rho_j(S\setminus\{j\})}{\rho_j(\emptyset)},$$

respectively. Assume without loss of generality that $F(\{j\}) > 0, \forall j \in \mathcal{V}$. One can observe that $\kappa_F(S) \leq \kappa_F$. A modular function has curvature $\kappa_F = 0$, and a matroid rank function has maximal curvature $\kappa_F = 1$. Vondrák (2010) also defines the relaxed notion of curvature (which is called curvature with respect to the optimum*) to be the smaller scalar $\bar\kappa_F(S)$ s.t,*

$$\rho_T(S) + \sum_{j\in S\cup T} \rho_j(S\cup T\setminus\{j\}) \geq (1-\bar\kappa_F(S))\rho_T(\emptyset), \forall T \subseteq \mathcal{V}. \tag{22}$$

Iyer et al. (2013) propose two new notions of curvature, which are,

$$\tilde\kappa_F(S) := 1 - \min_{T\subseteq\mathcal{V}} \frac{\rho_T(S) + \sum_{j\in S\cup T}\rho_j(S\cup T\setminus\{j\})}{\rho_T(\emptyset)},$$
$$\hat\kappa_F(S) := 1 - \frac{\sum_{j\in S}\rho_j(S\setminus\{j\})}{\sum_{j\in S}\rho_j(\emptyset)}.$$

Iyer et al. (2013) show that for submodular functions, it holds that $\hat\kappa_F(S) \leq \kappa_F(S) \leq \tilde\kappa_F(S) \leq \kappa_F$.

Submodularity ratio. Informally, the submodularity ratio quantifies how close a set function is to being submodular (Das & Kempe, 2011).

Definition 4 (Original submodularity ratio from Das & Kempe (2011))**.** *Let $F(\cdot)$ be a non-negative nondecreasing set function. The submodularity ratio of a set U w.r.t. an integer k is given by,*

$$\gamma_{U,k} := \min_{L\subseteq U} \min_{L,S: L\cap S=\emptyset, |S|\leq k} \frac{\sum_{j\in L}\rho_j(S)}{\rho_L(S)}.$$

C.2. Curvature of Non-submodular Functions and Relation to Our Results

Sviridenko et al. (2013) present a new notion of curvature for monotone set functions. We show how it is related to our notion of curvature in Def. 2. We also show that our approximation factors using the combination of curvature and submodularity ratio characterize the performance of GREEDY for solving problem (P) better.

Specifically, for a nondecreasing function F, Sviridenko et al. (2013, Section 8) define the curvature c as

$$1 - c = \min_{j\in\mathcal{V}} \min_{A,B\in\mathcal{V}\setminus\{j\}} \frac{\rho_j(A)}{\rho_j(B)}. \tag{23}$$

(Sviridenko et al., 2013, Theorem 8.1) show that for maximizing a nondecreasing function with bounded curvature $c \in [0, 1]$ under a matroid constraint, GREEDY enjoys an approximation guarantee of $(1 - c)$, and it is tight in terms of the definition of c in (23). The following remark discusses the relation to our definition of curvature.

Remark 3. *For a nondecreasing function $F(\cdot)$, it holds: a) c in (23) is always larger than the notion of curvature α in Def. 2, i.e., $c \geq \alpha$; b) For the* GREEDY *algorithm, there exists a class of functions for which the approximation guarantee characterized by c (which is $1 - c$) is strictly smaller than the approximation guarantee characterized by the combination of α and γ (which is $\alpha^{-1}(1 - e^{-\alpha\gamma})$ according to Theorem 1).*

Proof of Remark 3.

a) Note that the definition of curvature in Def. 2 is equivalent to the smallest scalar α such that,

$$\forall j \in \mathcal{V}, \forall B \subseteq A \in \mathcal{V}\setminus\{j\}, \rho_j(A) \geq (1-\alpha)\rho_j(B).$$

Now it is easy to see that $c \geq \alpha$.

b) Consider the class of functions in our tightness result in (3). From Lemma 4 we know that its curvature is α and submodularity ratio is γ. So its curvature c in (23) must be greater than or equal to α. Note that the approximation guarantee characterized by c is $1 - c \leq 1 - \alpha$. Taking $\alpha = 1$ in (3), the approximation guarantee of Sviridenko et al. (2013) is 0. While our approximation guarantee is γ, for any $\gamma \in (0, 1]$, our approximation guarantee is strictly higher than $1 - c$. □

C.3. Relation to Notions in This Work

- There are two versions of submodularity ratio in this paper: γ and γ^G, γ^G cannot be recovered from Def. 4. Our theory can easily accommodate Def. 4: our approximation guarantee in Theorem 1 holds for Def. 4 as long as U contains Ω^* and $k \geq K$. One benefit of the definition in this work (Def. 1) is that it better handles subtleties in Def. 4 where the denominator could be 0.

- The curvature in this work is a natural extension of the classical ones for monotone nondecreasing submodular functions (Conforti & Cornuéjols, 1984).

- Note that classical notions of curvature measure how close a submodular set function is to being modular. The notions of (generalized) curvature in Def. 2 measures how close a set function is to being *supermodular*.

- Our combinations of (generalized) curvature and submodularity ratio gives tight approximation guarantees for GREEDY, and this combination is more expressive than the curvature by Sviridenko et al. (2013), as shown in Remark 3.

D. Proofs for Bounding Parameters of Applications

D.1. Proving Proposition 1

Proof of Proposition 1.

Notice that in this subsection, the matrix $\mathbf{X}_S = [\boldsymbol{x}_{v_1}, \ldots, \boldsymbol{x}_{v_s}] \in \mathbb{R}^{d\times|S|}$ is the submatrix consisting the columns of $\mathbf{X}$ indexed by the set S.

Our proof considers the spectral parameters of the matrix $\mathbf{X}_S\mathbf{X}_S^\top$. For brevity, let us write $\mathbf{B} = \mathbf{\Lambda} + \sigma^{-2}\mathbf{X}_S\mathbf{X}_S^\top$. $\mathbf{B}$ is a symmetric positive definite matrix, thus can be factorized as $\mathbf{B} = \mathbf{PDP}^{-1}$.

Let the eigenvalues of $\mathbf{X}_S\mathbf{X}_S^\top$ be $\lambda_1(S) \geq \cdots \geq \lambda_d(S) \geq 0$, where we use the notation that $\lambda_i(S) := \lambda_i(\mathbf{X}_S\mathbf{X}_S^\top), \forall i \in [d]$. Then the eigenvalues of $\mathbf{B}$ are $\beta^2+\sigma^{-2}\lambda_i(S), i \in [d]$. One can see that $\mathbf{B}^{-1} = \mathbf{PD}^{-1}\mathbf{P}^{-1}$, and $\text{tr}(\mathbf{B}^{-1}) = \text{tr}(\mathbf{D}^{-1}) = \sum_{i=1}^d \frac{1}{\beta^2+\sigma^{-2}\lambda_i(S)}$.

Let the singular values of $\mathbf{X}_S$ be $\sigma_1(\mathbf{X}_S) \geq \cdots \geq \sigma_q(\mathbf{X}_S)$, where $q \leq \min\{d, |S|\}$. For notational simplicity, when $|S| < d$, we still use the convention $\sigma_i(\mathbf{X}_S) = 0, i = q+1, ..., d$ to represent the zeros values. One has $\sigma_i^2(\mathbf{X}_S) = \lambda_i(S), i = 1, ..., d$. For notational simplicity, we use $F(\cdot)$ to represent $F_A(\cdot)$ in the following.

Monotonicity. It can be easily seen that $F(\emptyset) = 0$. To prove that $F(S)$ is monotone nondecreasing, one just needs to show that $\forall \omega \in \mathcal{V} \setminus S$, it holds that $F(\{\omega\} \cup S) - F(S) \geq 0$. One can see that,

$$\begin{aligned} F(\{\omega\} \cup S) - F(S) &= \sum_{i=1}^d \frac{1}{\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_S)} - \sum_{j=1}^d \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_{S\cup\{\omega\}})} \\ &\geq 0 \quad \text{(Cauchy interlacing inequality of singular values)}. \end{aligned}$$

Bounding parameters. Let us restate the *assumption:* The data points are normalized, i.e., $\|\boldsymbol{x}_i\| = 1, \forall i \in \mathcal{V}$. Given this assumption, it holds that the spectral norm of the data matrix $\|\mathbf{X}\| = \sigma_{\max}(\mathbf{X}) \leq \sqrt{n}$, because of Weyl's inequality.

–Bounding the submodularity ratio: We need to lower bound $\frac{\sum_{\omega\in\Omega\setminus S}\rho_\omega(S)}{\rho_\Omega(S)} = \frac{\sum_{\omega\in\Omega\setminus S} F(\{\omega\}\cup S)-F(S)}{F(\Omega\cup S)-F(S)}$.

For the numerator, we have,

$$\begin{aligned} \sum_{\omega\in\Omega\setminus S} F(\{\omega\} \cup S) - F(S) &= \sum_{\omega\in\Omega\setminus S} \left[\sum_{i=1}^d \frac{1}{\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_S)} - \sum_{j=1}^d \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_{S\cup\{\omega\}})}\right] \\ &= \sum_{\omega\in\Omega\setminus S} \sum_{i=1}^d \frac{\sigma^{-2}[\sigma_i^2(\mathbf{X}_{S\cup\{\omega\}}) - \sigma_i^2(\mathbf{X}_S)]}{(\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_S))(\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_{S\cup\{\omega\}}))} \end{aligned}$$

$$
\begin{aligned}
&\geq (\beta^2 + \sigma^{-2}\sigma^2_{\max}(\mathbf{X}))^{-2} \sum_{\omega\in\Omega\setminus S}\sum_{i=1}^{d} \sigma^{-2}[\sigma_i^2(\mathbf{X}_{S\cup\{\omega\}}) - \sigma_i^2(\mathbf{X}_S)] \\
&= (\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \sum_{\omega\in\Omega\setminus S}\sum_{i=1}^{d} \sigma^{-2}[\lambda_i(S\cup\{\omega\}) - \lambda_i(S)] \\
&= (\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \sum_{\omega\in\Omega\setminus S} \sigma^{-2}[\mathrm{tr}(\mathbf{X}_{S\cup\{\omega\}}\mathbf{X}_{S\cup\{\omega\}}^\top) - \mathrm{tr}(\mathbf{X}_S\mathbf{X}_S^\top)] \\
&= (\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \sum_{\omega\in\Omega\setminus S} \sigma^{-2}[\mathrm{tr}(\mathbf{X}_S\mathbf{X}_S^\top + \boldsymbol{x}_\omega\boldsymbol{x}_\omega^\top) - \mathrm{tr}(\mathbf{X}_S\mathbf{X}_S^\top)] \\
&= (\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \sum_{\omega\in\Omega\setminus S} \sigma^{-2}\mathrm{tr}(\boldsymbol{x}_\omega\boldsymbol{x}_\omega^\top) \quad \text{(linearity of the trace)} \\
&= (\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \sum_{\omega\in\Omega\setminus S} \sigma^{-2}\|\boldsymbol{x}_\omega\|^2 \\
&= \sigma^{-2}(\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2}|\Omega\setminus S| \quad \text{(normalization of the data points)}
\end{aligned} \tag{24}
$$

For the denominator, one has,

$$
\begin{aligned}
F(\Omega\cup S) - F(S) &= \sum_{i=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_S)} - \sum_{j=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_{S\cup\Omega})} \\
&\leq \sum_{i=d-|\Omega\setminus S|+1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_i^2(\mathbf{X}_S)} - \sum_{j=1}^{|\Omega\setminus S|} \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_{S\cup\Omega})} \quad \text{(interlacing inequality of singular values)} \\
&\leq |\Omega\setminus S|(\frac{1}{\beta^2} - \frac{1}{\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2}) \\
&= |\Omega\setminus S| \frac{\sigma^{-2}\|\mathbf{X}\|^2}{\beta^2(\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)}.
\end{aligned} \tag{25}
$$

Combining (24) and (25) yields,

$$
\begin{aligned}
\frac{\sum_{\omega\in\Omega\setminus S} F(\{\omega\}\cup S) - F(S)}{F(\Omega\cup S) - F(S)} &\geq \frac{|\Omega\setminus S|\sigma^{-2}(\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2}}{|\Omega\setminus S|\frac{\sigma^{-2}\|\mathbf{X}\|^2}{\beta^2(\beta^2+\sigma^{-2}\|\mathbf{X}\|^2)}} \\
&= \frac{\beta^2}{\|\mathbf{X}\|^2(\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)}.
\end{aligned}
$$

–Bounding the curvature: We want to lower bound $1-\alpha$, which corresponds to lower bounding $\frac{F(S\cup\Omega) - F(S\setminus\{i\}\cup\Omega)}{F(S) - F(S\setminus\{i\})}$. For the numerator, one has,

$$
\begin{aligned}
F(S\cup\Omega) - F(S\setminus\{i\}\cup\Omega) &= \sum_{i'=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_{i'}^2(\mathbf{X}_{S\setminus\{i\}\cup\Omega})} - \sum_{j=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_{S\cup\Omega})} \\
&\geq \sigma^{-2}(\beta^2 + \sigma^{-2}\|\mathbf{X}\|^2)^{-2} \quad \text{(similar derivation as in (24)) .}
\end{aligned} \tag{26}
$$

For the denominator, one has (similar derivation as in (25)),

$$
\begin{aligned}
F(S) - F(S\setminus\{i\}) &= \sum_{i'=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_{i'}^2(\mathbf{X}_{S\setminus\{i\}})} - \sum_{j=1}^{d} \frac{1}{\beta^2 + \sigma^{-2}\sigma_j^2(\mathbf{X}_S)} \\
&\leq \frac{1}{\beta^2 + \sigma^{-2}\sigma_d^2(\mathbf{X}_{S\setminus\{i\}})} - \frac{1}{\beta^2 + \sigma^{-2}\sigma_1^2(\mathbf{X}_S)} \quad \text{(Cauchy interlacing inequality)}
\end{aligned}
$$

$$\leq \frac{\sigma^{-2}\|\mathbf{X}\|^2}{\beta^2(\beta^2+\sigma^{-2}\|\mathbf{X}\|^2)}. \tag{27}$$

Combining (26) and (27) we get,

$$\frac{F(S\cup\Omega)-F(S\setminus\{i\}\cup\Omega)}{F(S)-F(S\setminus\{i\})} \geq \frac{\beta^2}{\|\mathbf{X}\|^2(\beta^2+\sigma^{-2}\|\mathbf{X}\|^2)}.$$

□

D.2. Proofs for Determinantal Functions of Square Submatrix

Proof of Proposition 2.

Notice that in this subsection, the matrix $\mathbf{\Sigma}_S$ is the square submatrix of $\mathbf{\Sigma}$, with both its rows and columns indexed by S.

a) We want to prove that $F(\cdot)$ is supermodular. Assume that $A \subseteq B \subseteq \mathcal{V}$ and $i \in \mathcal{V}\setminus B$, then

$$\begin{aligned}
\rho_i(A) &= \det(\mathbf{I}+\sigma^{-2}\mathbf{\Sigma}_{A\cup\{i\}}) - \det(\mathbf{I}+\sigma^{-2}\mathbf{\Sigma}_A)\\
&= \sum_{S\subseteq A\cup\{i\}} \det((\sigma^{-2}\mathbf{\Sigma})_S) - \sum_{S\subseteq A}\det((\sigma^{-2}\mathbf{\Sigma})_S) \quad \text{(Kulesza \& Taskar, 2012, Theorem 2.1)}\\
&= \sum_{S\subseteq A}\det((\sigma^{-2}\mathbf{\Sigma})_{S\cup\{i\}})\\
&\leq \sum_{S\subseteq B}\det((\sigma^{-2}\mathbf{\Sigma})_{S\cup\{i\}}) \quad (\mathbf{\Sigma}\text{ is positive semidefinite})\\
&= \det(\mathbf{I}+\sigma^{-2}\mathbf{\Sigma}_{B\cup\{i\}}) - \det(\mathbf{I}+\sigma^{-2}\mathbf{\Sigma}_B)\\
&= \rho_i(B),
\end{aligned}$$

which proves that $F(\cdot)$ is supermodular.

b) We want to lower bound $\frac{\sum_{\omega\in\Omega\setminus S}\rho_\omega(S)}{\rho_\Omega(S)} = \frac{\sum_{\omega\in\Omega\setminus S}F(\{\omega\}\cup S)-F(S)}{F(\Omega\cup S)-F(S)}$.

For the numerator, one has,

$$\begin{aligned}
\sum_{\omega\in\Omega\setminus S} F(\{\omega\}\cup S)-F(S) &= \sum_{\omega\in\Omega\setminus S}\prod_{i=1}^{|S\cup\{\omega\}|}\lambda_i(\mathbf{A}_{S\cup\{\omega\}}) - \prod_{j=1}^{|S|}\lambda_j(\mathbf{A}_S)\\
&= \sum_{\omega\in\Omega\setminus S}\lambda_{|S\cup\{\omega\}|}(\mathbf{A}_{S\cup\{\omega\}})\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_{S\cup\{\omega\}}) - \prod_{j=1}^{|S|}\lambda_j(\mathbf{A}_S)\\
&\geq \sum_{\omega\in\Omega\setminus S}\lambda_{|S\cup\{\omega\}|}(\mathbf{A}_{S\cup\{\omega\}})\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S) - \prod_{j=1}^{|S|}\lambda_j(\mathbf{A}_S) \quad \text{(Cauchy interlacing inequality)}\\
&= \sum_{\omega\in\Omega\setminus S}(\lambda_{|S\cup\{\omega\}|}(\mathbf{A}_{S\cup\{\omega\}})-1)\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S).
\end{aligned} \tag{28}$$

For the denonimator, it holds,

$$\begin{aligned}
F(\Omega\cup S)-F(S) &= \prod_{i=|\Omega\setminus S|}^{|\Omega\cup S|}\lambda_i(\mathbf{A}_{\Omega\cup S})\prod_{j=1}^{|\Omega\setminus S|}\lambda_j(\mathbf{A}_{\Omega\cup S}) - \prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S)\\
&\leq \left(\prod_{j=1}^{|\Omega\setminus S|}\lambda_j(\mathbf{A}_{S\cup\Omega})-1\right)\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S) \quad \text{(Cauchy interlacing inequality).}
\end{aligned} \tag{29}$$

Combining (28) and (29) gives,

$$\begin{aligned}\frac{\sum_{\omega\in\Omega\setminus S} F(\{\omega\}\cup S) - F(S)}{F(\Omega\cup S) - F(S)} &\geq \frac{\sum_{\omega\in\Omega\setminus S}(\lambda_{|S\cup\{\omega\}|}(\mathbf{A}_{S\cup\{\omega\}})-1)\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S)}{\left(\prod_{j=1}^{|\Omega\setminus S|}\lambda_j(\mathbf{A}_{S\cup\Omega})-1\right)\prod_{i=1}^{|S|}\lambda_i(\mathbf{A}_S)} \\ &= \frac{\sum_{\omega\in\Omega\setminus S}(\lambda_{|S\cup\{\omega\}|}(\mathbf{A}_{S\cup\{\omega\}})-1)}{\left(\prod_{j=1}^{|\Omega\setminus S|}\lambda_j(\mathbf{A}_{S\cup\Omega})-1\right)} \\ &\geq \frac{K(\lambda_n-1)}{\prod_{j=1}^{K}\lambda_j - 1},\end{aligned}$$

where the last inequality comes from that $|\Omega\setminus S|\leq K$. □

D.3. LP with Combinatorial Constraints

D.3.1. Two examples where $F(S)$ is non-submodular

1), Considering the following LP:

$$\begin{array}{llll} \max & 4x_1+ & x_2+ & 4x_3 \\ \text{s.t.} & 2x_1+ & x_2 & \leq 2 \\ & & x_2+ & 2x_3 \leq 2 \\ & & x_1,x_2,x_3\geq 0. & \end{array} \tag{30}$$

For this LP, one can easily see that $F(\{1,2\}) = 4, F(\{2\}) = 2, F(\{1,2,3\}) = 8, F(\{2,3\}) = 4$, thus $F(\{1,2\}) - F(\{2\}) < F(\{1,2,3\}) - F(\{2,3\})$, which shows F is non-submodular.

2), Considering the following LP:

$$\begin{array}{lllll} \max & 10x_1+ & 12x_2+ & 12x_3 & \\ \text{s.t.} & x_1+ & 2x_2+ & 2x_3 & \leq 20 \\ & 2x_1+ & x_2+ & 2x_3 & \leq 20 \\ & 2x_2+ & 2x_2+ & x_3 & \leq 20 \\ & & & x_1,x_2,x_3\geq 0. & \end{array} \tag{31}$$

For this LP, one can see that $F(\{1,2\}) = 120, F(\{2\}) = 120, F(\{1,2,3\}) = 136, F(\{2,3\}) = 120$, thus $F(\{1,2\}) - F(\{2\}) < F(\{1,2,3\}) - F(\{2,3\})$. But this one has *degenerate* basic feasible solutions.

D.3.2. Proving Proposition 3

To prove Proposition 3, we first need to present the setup. The LP corresponding to $F(S)$ is,

$$(LP_S)\quad \begin{array}{ll} \max & \langle \boldsymbol{d}_S, \boldsymbol{x}_S\rangle \\ \text{s.t.} & \mathbf{A}_S\boldsymbol{x}_S \leq \boldsymbol{b} \\ & \boldsymbol{x}_S\geq 0. \end{array} \tag{32}$$

where the columns of $\mathbf{A}_S\in\mathbb{R}_+^{m\times|S|}$ are the columns of $\mathbf{A}$ indexed by the set S. $\boldsymbol{x}_S$ (respectively, $\boldsymbol{d}_S$) is the subvector of $\boldsymbol{x}$ (respectively, $\boldsymbol{d}$) indexed by S. To apply the optimality condition of a LP in the standard form, let us change (LP_S) to be the following standard LP by introducing the slack variable $\boldsymbol{\xi}\in\mathbb{R}^m$,

$$(LP_S^*)\quad \begin{array}{ll} -\min & \langle \boldsymbol{c}_S, \boldsymbol{x}_S\rangle \\ \text{s.t.} & \mathbf{A}_S\boldsymbol{x}_S + \mathbf{I}_m\boldsymbol{\xi} = \boldsymbol{b} \\ & \boldsymbol{x}_S\geq 0, \boldsymbol{\xi}\geq 0. \end{array} \tag{33}$$

where $\boldsymbol{c}_S := -\boldsymbol{d}_S$. Let us denote $\bar{\mathbf{A}} := [\mathbf{A}_S, \mathbf{I}_m]\in\mathbb{R}^{m\times(|S|+m)}$, $\bar{\boldsymbol{x}} := [\boldsymbol{x}_S^\top, \boldsymbol{\xi}^\top]^\top$.

Let $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ denote the optimal solution of (LP_S^*). The corresponding basis of of (LP_S^*) is $B^{(S)}$, which is a subset of $\mathcal{V} \cup \{\xi_1, \cdots, \xi_m\}$, and $|B^{(S)}| = m$.

According to Bertsimas & Tsitsiklis (1997, Chapter 3.1), the *optimality condition* for (LP_S^*) is: Given a basic feasible solution $(\boldsymbol{x}, \boldsymbol{\xi})$ with the basis as B, the reduced cost is $\bar{c}_j = c_j - \boldsymbol{c}_B^\top \bar{\mathbf{A}}_B^{-1} \bar{\mathbf{A}}_{\cdot j}$. 1) If $(\boldsymbol{x}, \boldsymbol{\xi})$ is optimal and non-degenerate, then $\bar{c}_j \geq 0, \forall j$; 2) If $\bar{c}_j \geq 0, \forall j$, then $(\boldsymbol{x}, \boldsymbol{\xi})$ is optimal.

Proof of Proposition 3. First of all, let us detail the non-degenerancy assumption.

Non-degenerancy assumption: The basic feasible solutions of the correpsonding LP in standard form (LP_S^*) is non-degenerate $\forall S \subseteq \mathcal{V}$.

a) It is easy to see that $F(\emptyset) = 0$, and $F(S)$ is nondecreasing.

b) For the submodularity ratio, we want to lower bound $\frac{\sum_{\omega \in \Omega \setminus S} \rho_\omega(S)}{\rho_\Omega(S)}$. There could be in total four situations:

1) $\sum_{\omega \in \Omega \setminus S} \rho_\omega(S) = 0$ but $\rho_\Omega(S) > 0$. We will prove that this situation cannot happen, or in the other words, $\sum_{\omega \in \Omega \setminus S} F(\{\omega\} \cup S) - F(S) = 0$ implies that $F(\Omega \cup S) - F(S) = 0$ as well.

First of all, since $F(S)$ is nondecreasing, so $F(\{\omega\} \cup S) - F(S) = 0, \forall \omega$. We know that $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ is the optimal solution of (LP_S^*), and $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ is a basic feasible solution of $(LP_{S\cup\{\omega\}}^*)$, so $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ is also the optimal solution of $(LP_{S\cup\{\omega\}}^*)$. Since $(LP_{S\cup\{\omega\}}^*)$ is non-degenerate, according to the optimality condition, the reduced cost of x_ω: $\bar{c}_\omega$ must be greater than or equal zero.

Now we know that $\bar{c}_\omega \geq 0, \forall \omega \in \Omega \setminus S$, and $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ is a basic feasible solution of $(LP_{S\cup\Omega}^*)$ as well, again using the optimality condition, we know that $(\boldsymbol{x}^{(S)}, \boldsymbol{\xi}^{(S)})$ is optimal for $(LP_{S\cup\{\Omega\}}^*)$. So $F(\Omega \cup S) - F(S) = 0$.

2) $\sum_{\omega \in \Omega \setminus S} \rho_\omega(S) = 0$ and $\rho_\Omega(S) = 0$. The submodularity ratio is 1 in this situation.

3) $\sum_{\omega \in \Omega \setminus S} \rho_\omega(S) > 0$ and $\rho_\Omega(S) = 0$. This can be ignored since we want a lower bound.

4) $\sum_{\omega \in \Omega \setminus S} \rho_\omega(S) > 0$ and $\rho_\Omega(S) > 0$. This situation gives the lower bound:

$$
\begin{aligned}
\frac{\sum_{\omega \in \Omega \setminus S} \rho_\omega(S)}{\rho_\Omega(S)} &\geq \frac{\max_{\omega \in \Omega \setminus S} \rho_\omega(S)}{F(\mathcal{V})} \\
&\geq \frac{\min_{S \subseteq \mathcal{V}, \omega \in \mathcal{V} \setminus S, \rho_\omega(S) > 0} \rho_\omega(S)}{F(\mathcal{V})} \\
&=: \gamma_0 > 0.
\end{aligned}
$$

□

E. Details about SDP Formulation of Bayesian A-optimality Objective

The SDP formulation used in this paper is consistent with that from Boyd & Vandenberghe (2004, Chapter 7.5) and Krause et al. (2008). To make this work self-contained, we present the details here.

Firstly, maximizing the Bayesian A-optimality objective is equivalent to,

$$
\min_{S \subseteq \mathcal{V}, |S| \leq K} \operatorname{tr}((\mathbf{\Lambda} + \sigma^{-2} \mathbf{X}_S \mathbf{X}_S^\top)^{-1}) \tag{34}
$$

By introducing binary variables $m_j, j \in [n]$, (34) is equivalent to,

$$
\begin{aligned}
&\min \operatorname{tr}((\mathbf{\Lambda} + \sigma^{-2} \sum_{j=1}^{n} m_j \boldsymbol{x}_j \boldsymbol{x}_j^\top)^{-1}) \qquad (35) \\
&\text{s.t. } m_j \in \{0, 1\}, j \in [n], m_1 + \cdots + m_n \leq K
\end{aligned}
$$

A proper relaxation is (relaxing the variables $\lambda_j = m_j/K, j \in [n]$),

$$\min \operatorname{tr}((\mathbf{\Lambda} + \sigma^{-2} \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j \boldsymbol{x}_j^\top)^{-1}) \tag{36}$$
$$\text{s.t. } \boldsymbol{\lambda} \in \mathbb{R}_+^n, \mathbf{1}^\top \boldsymbol{\lambda} = 1.$$

According to the Schur complement lemma, the relaxed formulation (36) is equivalent to the following SDP problem,

$$\min_{\boldsymbol{u} \in \mathbb{R}^d} \mathbf{1}^\top \boldsymbol{u}$$
$$\text{s.t. } \begin{bmatrix} \mathbf{\Lambda} + \sigma^{-2} \sum_{j=1}^{n} \lambda_j \boldsymbol{x}_j \boldsymbol{x}_j^\top & \boldsymbol{e}_k \\ \boldsymbol{e}_k^\top & u_k \end{bmatrix} \succeq 0, \quad k = 1, \cdots, d \quad \text{(SDP)}$$
$$\boldsymbol{\lambda} \in \mathbb{R}_+^n, \mathbf{1}^\top \boldsymbol{\lambda} = 1,$$

where $\boldsymbol{e}_k \in \mathbb{R}^d$ is the k^{th} standard basis vector. According to Krause et al. (2008), after solving the (SDP) problem we sort the entries of $\boldsymbol{\lambda}$ in descending order, and select the largest K coordinates as the indices of the K elements to be selected.

F. Proofs and Details in Related Work (Section 6)

Remark 4. *For a set function $F(\cdot)$: a) Its submodularity ratio γ is lower-bounded away from 0 and its curvature α is upper-bounded away from 1 does not imply that it is weakly submodular; b) $F(\cdot)$ is weakly submodular does not imply that its submodularity ratio γ is lower-bounded away from 0 and its curvature α is upper-bounded away from 1.*

Proof of Remark 4.

For argument a): Let $F(S) := |S|^4, S \subseteq \mathcal{V}$, which is a supermodular function, so the curvature is 0 (upper-bounded away from 1). The submodualrity ratio can be lower bounded by n^{-3}. But it is not weakly submodular according to Proposition 3.11 in Borodin et al. (2014).

For argument b): Let us take a minimum cardinality function with $k = 2$, i.e., $F(S) = B > 0$ iff. $|S| \geq 2$, otherwise $F(S) = 0$. According to Proposition 3.5 in Borodin et al. (2014), it is weakly submodular, but it is easy to see that its submodualrity ratio is 0. □

More on submodularity index. It is defined as (equivalent to that in Zhou & Spanos (2016)):

$$\min_{\Omega, S \subseteq \mathcal{V}} \min_{|\Omega \setminus S| \leq K} \Big(\sum_{\omega \in \Omega \setminus S} \rho_\omega(S) - \rho_\Omega(S) \Big).$$

G. More Applications

G.1. Subset Selection Using the R^2 Objective

Subset selection aims to estimate a predictor variable Z using linear regression on a small subset from the set of observation variables $\mathcal{V} = \{X_1, ..., X_n\}$. Let $\mathbf{C}$ to be the covariance matrix among the observation variables $\{X_1, ..., X_n\}$. We use $\boldsymbol{b}$ to denote the covariances between Z and the X_i, with entries $b_i = \text{Cov}(Z, X_i)$. Assuming there are m observations, let us arrange the data of all the observation variables to be a design matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, with each column representing the observations of one variable. Given a budget parameter K, subset selection tries to find a set $S \subseteq \mathcal{V}$ of at most K elements, and a linear predictor $Z' = \sum_{i \in S} \alpha_i X_i = \mathbf{X}_{\cdot S} \boldsymbol{\alpha}_S$, in order to maximize the squared multiple corrleation $R_{Z,S} = \frac{\text{Var}(Z) - \mathbb{E}[(Z - Z')^2]}{\text{Var}(Z)}$, it measures the fraction of variance of Z explained by variables in S. Assume Z is normalized to have variance 1, and it is well-known that the optimal regression coefficients are $\boldsymbol{\alpha}_S = (\mathbf{C}_S)^{-1} \boldsymbol{b}_S$, so the R^2 objective can be formulated as,

$$F(S) := R_{Z,S}^2 = \boldsymbol{b}_S^\top (\mathbf{C}_S)^{-1} \boldsymbol{b}_S, S \subseteq \mathcal{V}. \tag{37}$$

Das & Kempe (2011) show that the submodularity ratio of F in (37) can be lower bounded by $\lambda_{\min}(\mathbf{C})$, which is the smallest eigenvalue of $\mathbf{C}$. The theoretical results in this work suggests that the approximation guarantees for maximizing F in (37) can be further improved by analyzing the curvature parameters. The experimental results in Appendix H.2 demonstrates that it is promising to upper bound the curvature parameters of (37) (possibly with regular assumptions) .

G.2. Sparse Modeling with Strongly Convex Loss Functions

Sparse modeling aims to build a model with a small subset of at most K features, out of in total n features. Let $f(\boldsymbol{x}) : \mathbb{R}^n \mapsto \mathbb{R}$ to be the loss function, the corresponding objective is,

$$\min f(\boldsymbol{x}) \text{ s.t. } |\text{supp}(\boldsymbol{x})| \leq K.$$

Assume $f(\boldsymbol{x})$ is m-strongly convex and has Lipschitz continuous gradient with parameter L, which is equilavent to say that $g(\boldsymbol{x}) := -f(\boldsymbol{x})$ is m-strongly concave and has L-Lipschitz continuous gradient. Then for all $\boldsymbol{x}, \boldsymbol{y} \in \texttt{dom}(f)$ it holds,

$$\frac{m}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2 \leq -g(\boldsymbol{y}) + g(\boldsymbol{x}) + \langle \nabla g(\boldsymbol{x}), \boldsymbol{y} - \boldsymbol{x} \rangle \leq \frac{L}{2}\|\boldsymbol{y} - \boldsymbol{x}\|^2. \tag{38}$$

In solving this problem, the GREEDY algorithm maximizes the corresponding auxiliary set function,

$$F(S) := \max_{\text{supp}(\boldsymbol{x}) \subseteq S} g(\boldsymbol{x}), \; S \subseteq [n] \tag{39}$$

Elenberg et al. (2016) analyzed the approximation guarantees of GREEDY by bounding the submodularity ratio of $F(S)$. Specifically,

Lemma 5 (Paraphrasing Theorem 1 in Elenberg et al. (2016))**.** *The submodularity ratio of $F(S)$ in (39) is lower bounded by $\frac{m}{L}$.*

By further bounding the curvature parameters of the auxiliary set function in (39), one can get improved approximation guarantees according to our theoretical findings.

G.3. Optimal Budget Allocation with Combinatorial Constraints

Optimal budget allocation (Soma et al., 2014) is a special case of the influence maximization problem, it aims to distribute the budget (e.g., space of an inline advertisement, or time for a TV advertisement) among the customers, and to maximize the expected influence on the potential customers. A concrete application is for the *search marketing advertiser bidding* task, in which vendors bid for the right to appear alongside the results of different search keywords. Let $x_s^i \in \mathbb{R}_+$ to be the volume of advertising space allocated to the advertiser i to show his ad alongside query keyword s. Bian et al. (2017) present continuous DR-submodular objectives to model this problem with continuous assignments.

The search engine company (e.g., Google and Yahoo) needs to distribute the budget (ad space) to all vendors to maximize their influence on the customers, while respecting various continuous and *combinatorial* constraints. For the continuous constraints, for instance, each vendor has a specified budget limit for advertising, and the ad space associated with each search keyword can not be too large. These continuous constraints can be formulated as a convex set $\mathcal{P}$. For *combinatorial constraints*, each vendor needs to obey the Internet regulations of sensitive search keywords in his country, so the search engine company can only choose a subset of "legal" keywords for a specific vendor. The combinatorial constraints can be arranged as a matroid $\mathcal{M} = (\mathcal{V}, \mathcal{I})$. Hence the problem in general can be formulated as,

$$\max_{\boldsymbol{x} \in \mathcal{P} \text{ and } \text{supp}(\boldsymbol{x}) \in \mathcal{I}} g(\boldsymbol{x}),$$

where $g(\boldsymbol{x})$ is the total influence modeled by a DR-submodular function. For one of its possible forms, one can refer to Bian et al. (2017). The GREEDY algorithm solves this problem by maximizing the following auxiliary set function $F(S)$ while respecting the combinatorial constraints,

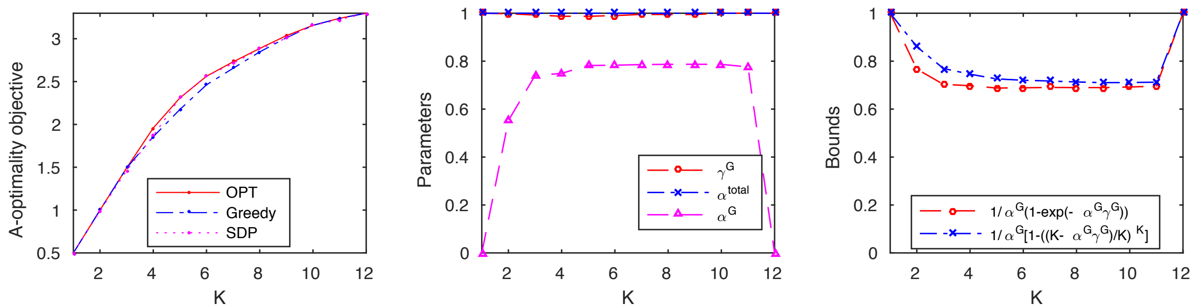
$$\max_{S \in \mathcal{I}} F(S), \text{ where } F(S) := \max_{\text{supp}(\boldsymbol{x}) \subseteq S, \boldsymbol{x} \in \mathcal{P}} g(\boldsymbol{x}). \tag{40}$$

By studying the submodularity ratio and curvature parameters of $F(S)$ in (40), one could obtain theoretical guarantees of the GREEDY algorithm according to Theorem 1 in this work.

H. More Experimental Results

H.1. Bayesian A-optimality Experiments

We put the results on a randomly generated dataset, to illustrate what does the proved bounds looks like. In the synthetic experiments we generate random observations from a multivariate Gaussian distribution with correlation 0.5. Fig. 8 shows



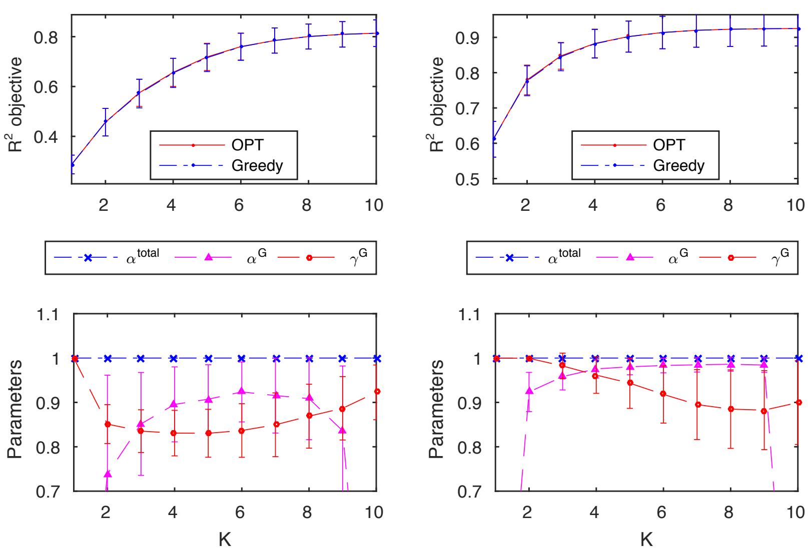
(a) A-optimality objective (b) Greedy submodularity ratio and curvature (c) Approximation bounds

Figure 8: Function value, parameters and approximation bounds of experimental design on synthetic data. Correlation: 0.5

the results (function value, parameters and approximation bounds) for one randomly generated data set with $d = 6$ features and $n = 12$ observations. Specifically, Fig. 8c traces the two approximation bounds from Theorem 1 (and Lemma 3): one curve shows the constant-factor bound $\alpha^{-1}(1 - e^{-\alpha\gamma})$ and the other the K-dependent bound $\frac{1}{\alpha}\left[1 - \left(\frac{K-\alpha\gamma}{K}\right)^K\right]$. We observe that both bounds give reasonable predictions of the performance of GREEDY.

H.2. Subset Selection Using the R^2 Objective

For details on this task please refer to Das & Kempe (2011) or Appendix G.1. We did synthetic experiments to illustrate that our theory can give a refined explanation of the performance of GREEDY. We generate random observations from a multivariate standard Gaussian distribution with different correlations. We used $n = 10$ features and $m = 100$ observations. The target regression coefficients $\boldsymbol{\alpha} \in \mathbb{R}^n$ were generated as a random vector with uniformly distributed entries in $[0, 1]$. Standard Gaussian noise was added to generate the observation of predictor variable Z. The results are shown in Fig. 9, with first column showing the results with correlation as 0.05, the second column with correlation as 0.5. One can see that the mean of the greedy curvature and submodularity ratio take values in $(0, 1)$, which can be used to give improved approximation bounds for GREEDY.



(a) Correlation: 0.05 (b) Correlation: 0.5

Figure 9: Results for R^2 objective on synthetic data.