Appendix

Robust Maximization of Non-Submodular Objectives

(Ilija Bogunovic†, Junyao Zhao† and Volkan Cevher, AISTATS 2018)

A Organization of the Appendix

– Appendix B: Proofs from Section 2
– Appendix C: Proofs of the Main Result (Section 3)
– Appendix D: Proofs from Section 4
– Appendix E: Additional experiments

B Proofs from Section 2

B.1 Proof of Proposition 1

Proof. We prove the following relations:

- $\boldsymbol{\nu \geq \gamma}$, $\boldsymbol{\check{\nu} \geq \check{\gamma}}$:
  By setting $S = \emptyset$ in both Eq. (4) and Eq. (5), we obtain $\forall S \subseteq V$:
$$\sum_{i \in S} f(\{i\}) \geq \gamma f(S), \tag{19}$$
  and
$$f(S) \geq \check{\gamma} \sum_{i \in S} f(\{i\}). \tag{20}$$
  The result follows since, by definition of ν and $\check{\nu}$, they are the largest scalars such that Eq. (19) and Eq. (20) hold, respectively.

- $\boldsymbol{\gamma \geq 1 - \check{\alpha}}$, $\boldsymbol{\check{\gamma} \geq 1 - \alpha}$:
  Let $S, \Omega \subseteq V$ be two arbitrary disjoint sets. We arbitrarily order elements of $\Omega = \{e_1, \cdots, e_{|\Omega|}\}$ and we let Ω_{j-1} denote the first $j-1$ elements of Ω. We also let Ω_0 be an empty set.

  By the definition of $\check{\alpha}$ (see Eq. (7)) we have:
$$\begin{aligned}\sum_{j=1}^{|\Omega|} f\left(\{e_j\}|S\right) &= \sum_{j=1}^{|\Omega|} f\left(\{e_j\}|S \cup \{e_j\} \setminus \{e_j\}\right) \\ &\geq \sum_{j=1}^{|\Omega|} (1-\check{\alpha}) f\left(\{e_j\}|S \cup \{e_j\} \setminus \{e_j\} \cup \Omega_{j-1}\right) \\ &= (1-\check{\alpha}) f\left(\Omega|S\right),\end{aligned} \tag{21}$$
  where the last equality is obtained via telescoping sums.

  Similarly, by the definition of α (see Eq. (6)) we have:
$$\begin{aligned}(1-\alpha)\sum_{j=1}^{|\Omega|} f\left(\{e_j\}|S\right) &= \sum_{j=1}^{|\Omega|} (1-\alpha) f\left(\{e_j\}|S \cup \{e_j\} \setminus \{e_j\}\right) \\ &\leq \sum_{j=1}^{|\Omega|} f\left(\{e_j\}|S \cup \{e_j\} \setminus \{e_j\} \cup \Omega_{j-1}\right) \\ &= f\left(\Omega|S\right).\end{aligned} \tag{22}$$

  Because S and Ω are arbitrary disjoint sets, and both γ and $\check{\gamma}$ are the largest scalars such that for all disjoint sets $S, \Omega \subseteq V$ the following holds $\sum_{j=1}^{|\Omega|} f(\{e_j\}|S) \geq \gamma f(\Omega|S)$ and $\check{\gamma} \sum_{j=1}^{|\Omega|} f\left(\{e_j\}|S\right) \leq f\left(\Omega|S\right)$, it follows from Eq. (21) and Eq. (22), respectively, that $\gamma \geq 1 - \check{\alpha}$ and $\check{\gamma} \geq 1 - \alpha$.

□

B.2 Proof of Remark 1

Proof. Consider any set $S \subseteq V$, and A and B such that $A \cup B = S$, $A \cap B = \emptyset$. We have

$$\frac{f(A)+f(B)}{f(S)} \geq \frac{\check{\nu}\sum_{i\in A} f(\{i\}) + \check{\nu}\sum_{i\in B} f(\{i\})}{f(S)} = \frac{\check{\nu}\sum_{i\in S} f(\{i\})}{f(S)} \geq \nu\check{\nu},$$

where the first and second inequality follow by the definition of ν and $\check{\nu}$ (Eq. (8) and Eq. (9)), respectively. By the definition (see Eq. (10)), θ is the largest scalar such that $f(A) + f(B) \geq \theta f(S)$ holds, hence, it follows $\theta \geq \nu\check{\nu}$. □

C Proofs of the Main Result (Section 3)

C.1 Proof of Lemma 2

We reproduce the proof from [2] for the sake of completeness.

Proof.

$$\begin{aligned}
f(S \setminus E_S^*) &= f(S) - f(S) + f(S \setminus E_S^*) \\
&= f(S_0 \cup S_1) + f(S \setminus E_0) - f(S \setminus E_0) - f(S) + f(S \setminus E_S^*) \\
&= f(S_1) + f(S_0 \mid S_1) + f(S \setminus E_0) - f(S) - f(S \setminus E_0) + f(S \setminus E_S^*) \\
&= f(S_1) + f(S_0 \mid (S \setminus S_0)) + f(S \setminus E_0) - f(E_0 \cup (S \setminus E_0)) - f(S \setminus E_0) + f(S \setminus E_S^*) \\
&= f(S_1) + f(S_0 \mid (S \setminus S_0)) - f(E_0 \mid (S \setminus E_0)) - f(S \setminus E_0) + f(S \setminus E_S^*) \\
&= f(S_1) + f(S_0 \mid (S \setminus S_0)) - f(E_0 \mid (S \setminus E_0)) - f(E_1 \cup (S \setminus E_S^*)) + f(S \setminus E_S^*) \\
&= f(S_1) + f(S_0 \mid (S \setminus S_0)) - f(E_0 \mid (S \setminus E_0)) - f(E_1 \mid S \setminus E_S^*) \\
&= f(S_1) - f(E_1 \mid S \setminus E_S^*) + f(S_0 \mid (S \setminus S_0)) - f(E_0 \mid (S \setminus E_0)) \\
&\geq (1-\mu) f(S_1), && (23)
\end{aligned}$$

where we used $S = S_0 \cup S_1$, $E_S^* = E_0 \cup E_1$. and (23) follows from monotonicity, i.e., $f(S_0 \mid (S \setminus S_0)) - f(E_0 \mid (S \setminus E_0)) \geq 0$ (due to $E_0 \subseteq S_0$ and $S \setminus S_0 \subseteq S \setminus E_0$), along with the definition of μ. □

C.2 Proof of Lemma 3

Proof. We start by defining $S_0' := \text{OPT}_{(k-\tau, V\setminus E_0)} \cap (S_0 \setminus E_0)$ and $X := \text{OPT}_{(k-\tau, V\setminus E_0)} \setminus S_0'$.

$$\begin{aligned}
f(S_0 \setminus E_0) + f(\text{OPT}_{(k-\tau, V\setminus S_0)}) &\geq f(S_0') + f(X) && (24)\\
&\geq \theta f(\text{OPT}_{(k-\tau, V\setminus E_0)}) && (25)\\
&\geq \theta f(\text{OPT}_{(k-\tau, V\setminus E_S^*)}), && (26)
\end{aligned}$$

where (24) follows from monotonicity as $S_0' \subseteq (S_0 \setminus E_0)$ and $(V \setminus S_0) \subseteq (V \setminus E_0)$. Eq. (25) follows from the fact that $\text{OPT}_{(k-\tau, V\setminus E_0)} = S_0' \cup X$ and the bipartite subadditive property (10). The final equation follows from the definition of the optimal solution and the fact that $E_S^* = E_0 \cup E_1$.

By rearranging and noting that $f(S \setminus E_S^*) \geq f(S_0 \setminus E_0)$ due to $(S_0 \setminus E_0) \subseteq (S \setminus E_S^*)$ and monotonicity, we obtain

$$f(S \setminus E_S^*) \geq \theta f(\text{OPT}_{(k-\tau, V\setminus E_S^*)}) - f(\text{OPT}_{(k-\tau, V\setminus S_0)}).$$

□

C.3 Proof of Theorem 1

Before proving the theorem we outline the following auxiliary lemma:

Lemma 5 (Lemma D.2 in [2])**.** *For any set function f, sets A, B, and constant $\alpha > 0$, we have*

$$\max\{\alpha f(A), \beta f(B) - f(A)\} \geq \left(\frac{\alpha}{1+\alpha}\right) \beta f(B). \tag{27}$$

Next, we prove the main theorem.

Proof. First we note that β should be chosen such that the following condition holds $|S_0| = \lceil \beta\tau \rceil \leq k$. When $\tau = \lceil ck \rceil$ for $c \in (0,1)$ and $k \to \infty$ the condition $\beta < \frac{1}{c}$ suffices.

We consider two cases, when $\mu = 0$ and $\mu \neq 0$. When $\mu = 0$, from Lemma 2 we have

$$f(S \setminus E_S^*) \geq f(S_1) \tag{28}$$

On the other hand, when $\mu \neq 0$, by Lemma 2 and 4 we have

$$\begin{aligned} f(S \setminus E_S^*) &\geq \max\{(1-\mu) f(S_1), (\beta - 1)\check{\nu}(1-\check{\alpha})\mu f(S_1)\} \\ &\geq \frac{(\beta-1)\check{\nu}(1-\check{\alpha})}{1 + (\beta-1)\check{\nu}(1-\check{\alpha})} f(S_1). \end{aligned} \tag{29}$$

By denoting $P := \frac{(\beta-1)\check{\nu}(1-\check{\alpha})}{1+(\beta-1)\check{\nu}(1-\check{\alpha})}$ we observe that $P \in [0,1)$ once $\beta \geq 1$. Hence, by setting $\beta \geq 1$ and taking the minimum between two bounds in Eq. (29) and Eq. (28) we conclude that Eq. (29) holds for any $\mu \in [0,1]$.

By combining Eq. (29) with Lemma 1 we obtain

$$f(S \setminus E_S^*) \geq P\left(1 - e^{-\gamma \frac{k - \lceil \beta\tau \rceil}{k-\tau}}\right) f(\mathrm{OPT}_{(k-\tau, V \setminus S_0)}). \tag{30}$$

By further combining this with Lemma 3 we have

$$\begin{aligned} f(S \setminus E_S^*) &\geq \max\{\theta f(\mathrm{OPT}_{(k-\tau, V\setminus E_S^*)}) - f(\mathrm{OPT}_{(k-\tau, V\setminus S_0)}), P\left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right) f(\mathrm{OPT}_{(k-\tau, V\setminus S_0)})\} \\ &\geq \theta \frac{P\left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right)}{1 + P\left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right)} f(\mathrm{OPT}_{(k-\tau, V\setminus E_S^*)}) \end{aligned} \tag{31}$$

where the second inequality follows from Lemma 5. By plugging in $\tau = \lceil ck \rceil$ we further obtain

$$\begin{aligned} f(S \setminus E_S^*) &\geq \theta \frac{P\left(1 - e^{-\gamma \frac{k - \beta\lceil ck \rceil - 1}{(1-c)k}}\right)}{1 + P\left(1 - e^{-\gamma \frac{k - \beta\lceil ck \rceil - 1}{(1-c)k}}\right)} f(\mathrm{OPT}_{(k-\tau, V\setminus E_S^*)}) \\ &\geq \theta \frac{P\left(1 - e^{-\gamma \frac{1 - \beta c - \frac{1}{k} - \frac{\beta}{k}}{1-c}}\right)}{1 + P\left(1 - e^{-\gamma \frac{1 - \beta c - \frac{1}{k} - \frac{\beta}{k}}{1-c}}\right)} f(\mathrm{OPT}_{(k-\tau, V\setminus E_S^*)}) \\ &\xrightarrow{k\to\infty} \frac{\theta P\left(1 - e^{-\gamma \frac{1-\beta c}{1-c}}\right)}{1 + P\left(1 - e^{-\gamma \frac{1-\beta c}{1-c}}\right)} f(\mathrm{OPT}_{(k-\tau, V\setminus E_S^*)}). \end{aligned}$$

Finally, Remark 2 follows from Eq. (30) when $\tau \in o\left(\frac{k}{\beta}\right)$ and $\beta \geq \log k$ (note that the condition $|S_0| = \lceil \beta\tau \rceil \leq k$ is thus satisfied), as $k \to \infty$, we have both $\frac{k - \lceil \beta\tau \rceil}{k-\tau} \to 1$ and $P = \frac{(\beta-1)\check{\nu}(1-\check{\alpha})}{1+(\beta-1)\check{\nu}(1-\check{\alpha})} \to 1$, when $\check{\nu} \in (0,1]$ and $\check{\alpha} \in [0,1)$.

$\square$

C.4 Proof of Corollary 1

To prove this result we need the following two lemmas that can be thought of as the alternative to Lemma 2 and 4.

Lemma 6. *Let $\mu' \in [0,1]$ be a constant such that $f(E_1) = \mu' f(S_1)$ holds. Consider $f(\cdot)$ with bipartite subadditivity ratio $\theta \in [0,1]$ defined in Eq. (4). Then*

$$f(S \setminus E_S^*) \geq (\theta - \mu') f(S_1). \tag{32}$$

Proof. By the definition of θ, $f(S_1 \setminus E_1) + f(E_1) \geq \theta f(S_1)$. Hence,

$$\begin{aligned} f(S \setminus E_S^*) &\geq f(S_1 \setminus E_1) \\ &\geq \theta f(S_1) - f(E_1) \\ &= (\theta - \mu') f(S_1). \end{aligned}$$

□

Lemma 7. *Let β be a constant such that $|S_0| = \lceil \beta\tau \rceil$ and $|S_0| \leq k$, and let $\check{\nu}, \nu \in [0,1]$ be superadditivity and subadditivity ratio (Eq. (9) and Eq. (8), respectively). Finally, let μ' be a constant defined as in Lemma 6. Then,*

$$f(S \setminus E_S^*) \geq (\beta - 1)\check{\nu}\nu\mu' f(S_1). \tag{33}$$

Proof. The proof follows that of Lemma 4, with two modifications. In Eq. (34) we used the subadditive property of $f(\cdot)$, and Eq. (35) follows by the definition of μ'.

$$\begin{aligned} f(S \setminus E_S^*) &\geq f(S_0 \setminus E_0) \\ &\geq \check{\nu} \sum_{e_i \in S_0 \setminus E_0} f(\{e_i\}) \\ &\geq \frac{|S_0 \setminus E_0|}{|E_1|} \check{\nu} \sum_{e_i \in E_1} f(\{e_i\}) \\ &\geq \frac{(\beta - 1)\tau}{\tau} \check{\nu} \sum_{e_i \in E_1} f(\{e_i\}) \\ &\geq (\beta - 1)\check{\nu}\nu f(E_1) \qquad (34) \\ &= (\beta - 1)\check{\nu}\nu\mu' f(S_1). \qquad (35) \end{aligned}$$

□

Next we prove the main corollary. The proof follows the steps of the proof from Appendix C.3, except that here we make use of Lemma 6 and 7.

Proof. We consider two cases, when $\mu' = 0$ and $\mu' \neq 0$. When $\mu' = 0$, from Lemma 6 we have

$$f(S \setminus E_S^*) \geq \theta f(S_1).$$

On the other hand, when $\mu' \neq 0$, by Lemma 6 and 7 we have

$$\begin{aligned} f(S \setminus E_S^*) &\geq \max\{(\theta - \mu') f(S_1), (\beta - 1)\check{\nu}\nu\mu' f(S_1)\} \\ &\geq \theta \frac{(\beta - 1)\check{\nu}\nu}{1 + (\beta - 1)\check{\nu}\nu} f(S_1). \qquad (36) \end{aligned}$$

By denoting $P := \frac{(\beta-1)\check{\nu}\nu}{1+(\beta-1)\check{\nu}\nu}$ and observing that $P \in [0,1)$ once $\beta \geq 1$, we conclude that Eq. (36) holds for any $\mu' \in [0,1]$ once $\beta \geq 1$.

By combining Eq. (36) with Lemma 1 we obtain

$$f(S \setminus E_S^*) \geq \theta P \left(1 - e^{-\gamma \frac{k - \lceil \beta\tau \rceil}{k - \tau}}\right) f(\mathrm{OPT}_{(k-\tau, V \setminus S_0)}). \tag{37}$$

By further combining this with Lemma 3 we have

$$f(S \setminus E_S^*) \geq \max\{\theta f(\text{OPT}_{(k-\tau, V\setminus E_S^*)}) - f(\text{OPT}_{(k-\tau,V\setminus S_0)}), \theta P \left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right) f(\text{OPT}_{(k-\tau,V\setminus S_0)})\}$$
$$\geq \frac{\theta^2 P \left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right)}{1 + \theta P \left(1 - e^{-\gamma \frac{k-\lceil \beta\tau \rceil}{k-\tau}}\right)} f(\text{OPT}_{(k-\tau,V\setminus E_S^*)}), \tag{38}$$

where the second inequality follows from Lemma 5. By plugging in $\tau = \lceil ck \rceil$ in the last equation and by letting $k \to \infty$ we arrive at:

$$f(S \setminus E_S^*) \geq \frac{\theta^2 P \left(1 - e^{-\gamma \frac{1-\beta c}{1-c}}\right)}{1 + \theta P \left(1 - e^{-\gamma \frac{1-\beta c}{1-c}}\right)} f(\text{OPT}_{(k-\tau,V\setminus E_S^*)}).$$

Finally, from Eq. (38), when $\tau \in o\left(\frac{k}{\beta}\right)$ and $\beta \geq \log k$, as $k \to \infty$, we have both $\frac{k-\lceil \beta\tau \rceil}{k-\tau} \to 1$ and $P = \frac{(\beta-1)\check{\nu}\nu}{1+(\beta-1)\check{\nu}\nu} \to 1$ (when $\nu, \check{\nu} \in (0,1]$). It follows

$$f(S \setminus E_S^*) \xrightarrow{k\to\infty} \frac{\theta^2(1-e^{-\gamma})}{1+\theta(1-e^{-\gamma})} f(\text{OPT}_{(k-\tau,V\setminus E_S^*)}).$$

□

D Proofs from Section 4

D.1 Proof of Proposition 2

Proof. The goal is to prove: $\check{\gamma} \geq \frac{m}{L}$.

Let $S \subseteq [d]$ and $\Omega \subseteq [d]$ be any two disjoint sets, and for any set $A \subseteq [d]$ let $\mathbf{x}^{(A)} = \arg\max_{\text{supp}(\mathbf{x})\subseteq A, \mathbf{x}\in\mathcal{X}} l(\mathbf{x})$. Moreover, for $B \subseteq [d]$ let $\mathbf{x}_B^{(A)}$ denote those coordinates of vector $\mathbf{x}^{(A)}$ that correspond to the indices in B.

We proceed by upper bounding the denominator and lower bounding the numerator in (5). By definition of $\mathbf{x}^{(S)}$ and strong concavity of $l(\cdot)$,

$$\begin{aligned} l(\mathbf{x}^{(S\cup\{i\})}) - l(\mathbf{x}^{(S)}) &\leq \langle \nabla l(\mathbf{x}^{(S)}), \mathbf{x}^{(S\cup\{i\})} - \mathbf{x}^{(S)} \rangle - \frac{m}{2} \left\| \mathbf{x}^{(S\cup\{i\})} - \mathbf{x}^{(S)} \right\|^2 \\ &\leq \max_{\mathbf{v}:\mathbf{v}_{(S\cup\{i\})^c}=0} \langle \nabla l(\mathbf{x}^{(S)}), \mathbf{v} - \mathbf{x}^{(S)} \rangle - \frac{m}{2} \left\| \mathbf{v} - \mathbf{x}^{(S)} \right\|^2 \\ &= \frac{1}{2m} \left\| \nabla l(\mathbf{x}^{(S)})_i \right\|^2 \end{aligned}$$

where the last equality follows by plugging in the maximizer $\mathbf{v} = \mathbf{x}^{(S)} + \frac{1}{m}\nabla l(\mathbf{x}^{(S)})_i$. Hence,

$$\sum_{i\in\Omega} \left(l(\mathbf{x}^{(S\cup\{i\})}) - l(\mathbf{x}^{(S)}) \right) \leq \sum_{i\in\Omega} \frac{1}{2m} \left\| \nabla l(\mathbf{x}^{(S)})_i \right\|^2 = \frac{1}{2m} \left\| \nabla l(\mathbf{x}^{(S)})_\Omega \right\|^2.$$

On the other hand, from the definition of $\mathbf{x}^{(S\cup\Omega)}$ and due to smoothness of $l(\cdot)$ we have

$$\begin{aligned} l(\mathbf{x}^{(S\cup\Omega)}) - l(\mathbf{x}^{(S)}) &\geq l(\mathbf{x}^{(S)} + \frac{1}{L}\nabla l(\mathbf{x}^{(S)})_\Omega) - l(\mathbf{x}^{(S)}) \\ &\geq \langle \nabla l(\mathbf{x}^{(S)}), \frac{1}{L}\nabla l(\mathbf{x}^{(S)})_\Omega \rangle - \frac{L}{2} \left\| \frac{1}{L}\nabla l(\mathbf{x}^{(S)})_\Omega \right\|^2 \\ &= \frac{1}{2L} \left\| l(\mathbf{x}^{(S)})_\Omega \right\|^2. \end{aligned}$$

It follows that

$$\frac{l(\mathbf{x}^{(S\cup\Omega)}) - l(\mathbf{x}^{(S)})}{\sum_{i\in\Omega}\left(l(\mathbf{x}^{(S\cup\{i\})}) - l(\mathbf{x}^{(S)})\right)} \geq \frac{m}{L}, \quad \forall \text{ disjoint } S, \Omega \subseteq [d]$$

We finish the proof by noting that $\check{\gamma}$ is the largest constant for the above statement to hold.

□

D.2 Variance Reduction in GPs

D.2.1 Non-submodularity of Variance Reduction

The goal of this section is to show that the GP variance reduction objective is not submodular in general. Consider the following PSD kernel matrix:

$$\mathbf{K} = \begin{bmatrix} 1 & \sqrt{1-z^2} & 0 \\ \sqrt{1-z^2} & 1 & z^2 \\ 0 & z^2 & 1 \end{bmatrix}.$$

We consider a single $x = \{3\}$ (i.e. M is a singleton) that corresponds to the third data point. The objective is as follows:

$$F(i|S) = \sigma^2_{\{3\}|S} - \sigma^2_{\{3\}|S\cup i}.$$

The submodular property implies $F(\{1\}) \geq F(\{1\}|\{2\})$. We have:

$$\begin{aligned} F(\{1\}) &= \sigma^2_{\{3\}} - \sigma^2_{\{3\}|\{1\}} \\ &= 1 - K(\{3\},\{3\}) - K(\{3\},\{1\})(K(\{1\},\{1\}) + \sigma^2)^{-1}K(\{1\},\{3\}) \\ &= 1 - 1 + 0 = 0, \end{aligned}$$

and

$$\begin{aligned} F(\{2\}) &= \sigma^2_{\{3\}} - \sigma^2_{\{3\}|\{2\}} \\ &= 1 - K(\{3\},\{3\}) - K(\{3\},\{2\})(K(\{2\},\{2\}) + \sigma^2)^{-1}K(\{2\},\{3\}) \\ &= 1 - (1 - z^2(1+\sigma^2)^{-1}z^2) = \frac{z^4}{1+\sigma^2}, \end{aligned}$$

and

$$\begin{aligned} F(\{1,2\}) &= \sigma^2_{\{3\}} - \sigma^2_{\{3\}|\{1,2\}} \\ &= 1 - K(\{3\},\{3\}) + [K(\{3\},\{1\}), K(\{3\},\{2\})] \begin{bmatrix} 1+\sigma^2, K(\{2\},\{1\}) \\ K(\{1\},\{2\}), 1+\sigma^2 \end{bmatrix}^{-1} \begin{bmatrix} K(\{1\},\{3\}) \\ K(\{2\},\{3\}) \end{bmatrix} \\ &= 1 - 1 + [0, z^2] \begin{bmatrix} 1+\sigma^2, \sqrt{1-z^2} \\ \sqrt{1-z^2}, 1+\sigma^2 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ z^2 \end{bmatrix} \\ &= \frac{z^4(1+\sigma^2)}{(1+\sigma^2)^2 - (1-z^2)}. \end{aligned}$$

We obtain,

$$\begin{aligned} F(\{1\}|\{2\}) &= F(\{1,2\}) - F(\{2\}) \\ &= \frac{z^4}{(1+\sigma^2) - (1-z^2)(1+\sigma^2)^{-1}} - \frac{z^4}{1+\sigma^2}. \end{aligned}$$

When $z \in (0,1)$, $F(\{1\}|\{2\})$ is strictly greater than 0, and hence greater than $F(\{1\})$. This is in contradiction with the submodular property which implies $F(\{1\}) \geq F(\{1\}|\{2\})$.

D.2.2 Proof of Proposition 3

Proof. We are interested in lower bounding the following ratios: $\frac{f(\{i\}|S\setminus\{i\}\cup\Omega)}{f(\{i\}|S\setminus\{i\})}$ and $\frac{f(\{i\}|S\setminus\{i\})}{f(\{i\}|S\setminus\{i\}\cup\Omega)}$.

Let $k_{\max} \in \mathbb{R}_+$ be the largest variance, i.e. $k(\mathbf{x}_i, \mathbf{x}_i) \le k_{\max}$ for every i. Consider the case when M is a singleton set:

$$f(i|S) = \sigma^2_{\mathbf{x}|S} - \sigma^2_{\mathbf{x}|S\cup i}.$$

By using $\Omega = \{i\}$ in Eq. (39), we can rewrite $f(i|S)$ as

$$f(i|S) = a_i^2 B_i^{-1},$$

where $a_i, B_i \in \mathbb{R}_+$, and are given by:

$$a_i = k(\mathbf{x}, \mathbf{x}_i) - k(\mathbf{x}, \mathbf{X}_S)(k(\mathbf{X}_S, \mathbf{X}_S) + \sigma^2\mathbf{I})^{-1} k(\mathbf{X}_S, \mathbf{x}_i)$$

and

$$B_i = \sigma^2 + k(\mathbf{x}_i, \mathbf{x}_i) - k(\mathbf{x}_i, \mathbf{X}_S)(k(\mathbf{X}_S, \mathbf{X}_S) + \sigma^2\mathbf{I})^{-1} k(\mathbf{X}_S, \mathbf{x}_i).$$

By using the fact that $k(\mathbf{x}_i, \mathbf{x}_i) \le k_{\max}$, for every i and S, we can upper bound B_i by $\sigma^2 + k_{\max}$ (note that $k(\mathbf{x}_i, \mathbf{x}_i) - k(\mathbf{x}_i, \mathbf{X}_S)(k(\mathbf{X}_S, \mathbf{X}_S) + \sigma^2\mathbf{I})^{-1} k(\mathbf{X}_S, \mathbf{x}_i) \ge 0$ as variance cannot be negative), and lower bound by σ^2. It follows that for every i and S we have:

$$\frac{a_i^2}{\sigma^2 + k_{\max}} \le f(i|S) \le \frac{a_i^2}{\sigma^2}.$$

Therefore,

$$\frac{f(\{i\}|S\setminus\{i\}\cup\Omega)}{f(\{i\}|S\setminus\{i\})} \ge \frac{a_i^2/(\sigma^2 + k_{\max})}{a_i^2/\sigma^2} = \frac{\sigma^2}{\sigma^2 + k_{\max}}, \quad \forall S, \Omega \subseteq V, i \in S\setminus\Omega,$$
$$\frac{f(\{i\}|S\setminus\{i\})}{f(\{i\}|S\setminus\{i\}\cup\Omega)} \ge \frac{a_i^2/(\sigma^2 + k_{\max})}{a_i^2/\sigma^2} = \frac{\sigma^2}{\sigma^2 + k_{\max}}, \quad \forall S, \Omega \subseteq V, i \in S\setminus\Omega.$$

It follows:

$$(1-\alpha) \ge \frac{\sigma^2}{\sigma^2 + k_{\max}}, \quad \text{and}$$
$$(1-\breve{\alpha}) \ge \frac{\sigma^2}{\sigma^2 + k_{\max}}.$$

The obtained result also holds for any set $M \subseteq [n]$. ☐

D.2.3 Alternative GP variance reduction form

Here, the goal is to show that the variance reduction can be written as

$$F(\Omega|S) = \sigma^2_{\mathbf{x}|S} - \sigma^2_{\mathbf{x}|S\cup\Omega} = \mathbf{a}\mathbf{B}^{-1}\mathbf{a}^T, \tag{39}$$

where $\mathbf{a} \in \mathbb{R}_+^{1\times|\Omega\setminus S|}$, $\mathbf{B} \in \mathbb{R}_+^{|\Omega\setminus S|\times|\Omega\setminus S|}$ and are given by:

$$\mathbf{a} := k(\mathbf{x}, \mathbf{X}_{\Omega\setminus S}) - k(\mathbf{x}, \mathbf{X}_S)(k(\mathbf{X}_S, \mathbf{X}_S) + \sigma^2\mathbf{I})^{-1} k(\mathbf{X}_S, \mathbf{X}_{\Omega\setminus S}),$$

and

$$\mathbf{B} := \sigma^2\mathbf{I} + k(\mathbf{X}_{\Omega\setminus S}, \mathbf{X}_{\Omega\setminus S}) - k(\mathbf{X}_{\Omega\setminus S}, \mathbf{X}_S)(k(\mathbf{X}_S, \mathbf{X}_S) + \sigma^2\mathbf{I})^{-1} k(\mathbf{X}_S, \mathbf{X}_{\Omega\setminus S}).$$

This form is used in the proof in Appendix D.2.2.

Proof. Recall the definition of the posterior variance:

$$\sigma^2_{\mathbf{x}|S} = k(\mathbf{x},\mathbf{x}) - k(\mathbf{x},\mathbf{X}_S)\left(k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I}_{|S|}\right)^{-1} k(\mathbf{X}_S,\mathbf{x}).$$

We have

$$\begin{aligned}
F(\Omega|S) &= \sigma^2_{\mathbf{x}|S} - \sigma^2_{\mathbf{x}|S\cup\Omega} \\
&= k(\mathbf{x},\mathbf{X}_{S\cup\Omega})\left(k(\mathbf{X}_{S\cup\Omega},\mathbf{X}_{S\cup\Omega}) + \sigma^2\mathbf{I}_{|\Omega\cup S|}\right)^{-1} k(\mathbf{X}_{S\cup\Omega},\mathbf{x}) - k(\mathbf{x},\mathbf{X}_S)\left(k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I}_{|S|}\right)^{-1} k(\mathbf{X}_S,\mathbf{x}) \\
&= [\mathbf{m}_1,\mathbf{m}_2]\begin{bmatrix}\mathbf{A}_{11},\mathbf{A}_{12}\\ \mathbf{A}_{21},\mathbf{A}_{22}\end{bmatrix}^{-1}\begin{bmatrix}\mathbf{m}_1^T\\ \mathbf{m}_2^T\end{bmatrix} - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{m}_1^T,
\end{aligned}$$

where we use the following notation:

$$\begin{aligned}
\mathbf{m}_1 &:= k(\mathbf{x},\mathbf{X}_S),\\
\mathbf{m}_2 &:= k(\mathbf{x},\mathbf{X}_{\Omega\setminus S}),\\
\mathbf{A}_{11} &:= k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I}_{|S|},\\
\mathbf{A}_{12} &:= k(\mathbf{X}_S,\mathbf{X}_{\Omega\setminus S}),\\
\mathbf{A}_{21} &:= k(\mathbf{X}_{\Omega\setminus S},X_S),\\
\mathbf{A}_{22} &:= k(\mathbf{X}_{\Omega\setminus S},\mathbf{X}_{\Omega\setminus S}) + \sigma^2\mathbf{I}_{|\Omega\setminus S|}.
\end{aligned}$$

By using the inverse formula [39, Section 9.1.3] we obtain:

$$F(\Omega|S) = [\mathbf{m}_1,\mathbf{m}_2]\begin{bmatrix}\mathbf{A}_{11}^{-1} + \mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}, & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\\ -\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}, & \mathbf{B}^{-1}\end{bmatrix}\begin{bmatrix}\mathbf{m}_1^T\\ \mathbf{m}_2^T\end{bmatrix} - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{m}_1^T,$$

where

$$\mathbf{B} := \mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}.$$

Finally, we obtain:

$$\begin{aligned}
F(\Omega|S) &= \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{m}_1^T + \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T - \mathbf{m}_2\mathbf{B}^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T\\
&\quad - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}\mathbf{m}_2^T + \mathbf{m}_2\mathbf{B}^{-1}\mathbf{m}_2^T - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{m}_1^T\\
&= \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\mathbf{B}^{-1}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T - \mathbf{m}_2^T) - \mathbf{m}_2\mathbf{B}^{-1}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T - \mathbf{m}_2^T)\\
&= (\mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12} - \mathbf{m}_2)\mathbf{B}^{-1}(\mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T - \mathbf{m}_2^T)\\
&= (\mathbf{m}_2 - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12})\mathbf{B}^{-1}(\mathbf{m}_2^T - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T).
\end{aligned}$$

By setting

$$\begin{aligned}
\mathbf{a} &:= \mathbf{m}_2 - \mathbf{m}_1\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\\
&= k(\mathbf{x},\mathbf{X}_{\Omega\setminus S}) - k(\mathbf{x},\mathbf{X}_S)(k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I})^{-1}k(\mathbf{X}_S,\mathbf{X}_{\Omega\setminus S})
\end{aligned}$$

and

$$\begin{aligned}
\mathbf{a}^T &:= \mathbf{m}_2^T - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{m}_1^T\\
&= k(\mathbf{X}_{\Omega\setminus S},\mathbf{x}) - k(\mathbf{X}_{\Omega\setminus S},\mathbf{X}_S)(k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I})^{-1}k(\mathbf{X}_S,\mathbf{x}),
\end{aligned}$$

we have

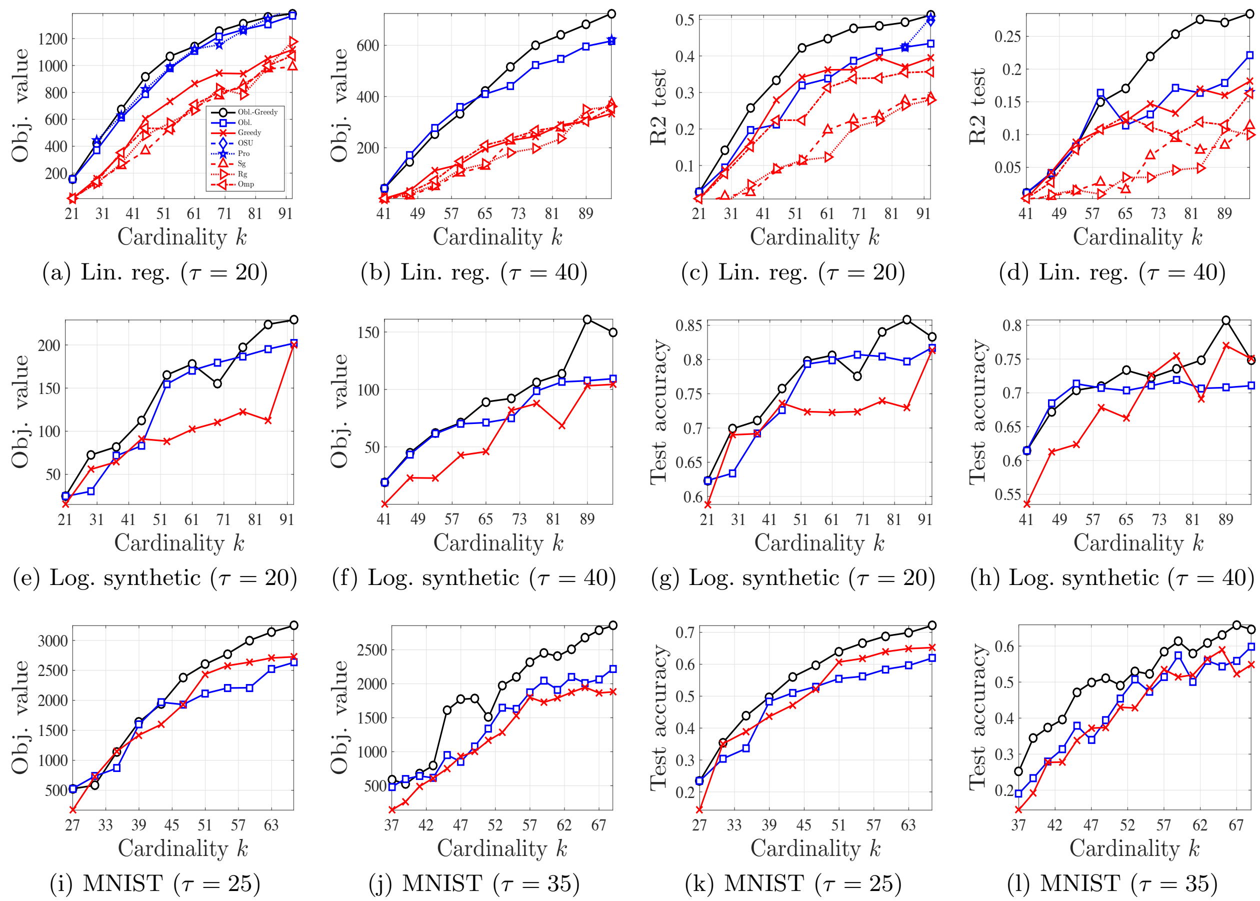
$$F(\Omega|S) = \mathbf{a}\mathbf{B}^{-1}\mathbf{a}^T,$$

where

$$\mathbf{B} = \sigma^2\mathbf{I}_{|\Omega\setminus S|} + k(\mathbf{X}_{\Omega\setminus S},\mathbf{X}_{\Omega\setminus S}) - k(\mathbf{X}_{\Omega\setminus S},\mathbf{X}_S)(k(\mathbf{X}_S,\mathbf{X}_S) + \sigma^2\mathbf{I}_{|S|})^{-1}k(\mathbf{X}_S,\mathbf{X}_{\Omega\setminus S}).$$

□

E Additional Experiments



(a) Lin. reg. ($\tau = 20$) (b) Lin. reg. ($\tau = 40$) (c) Lin. reg. ($\tau = 20$) (d) Lin. reg. ($\tau = 40$)

(e) Log. synthetic ($\tau = 20$) (f) Log. synthetic ($\tau = 40$) (g) Log. synthetic ($\tau = 20$) (h) Log. synthetic ($\tau = 40$)

(i) MNIST ($\tau = 25$) (j) MNIST ($\tau = 35$) (k) MNIST ($\tau = 25$) (l) MNIST ($\tau = 35$)

Figure 6: Additional experiments for comparison of the algorithms on support selection task.