Supplementary Material for "On the Unreasonable Effectiveness of the Greedy Algorithm: Greedy Adapts to Sharpness"

Sebastian Pokutta [1] **Mohit Singh** [2] **Alfredo Torrico** [3]

A. Appendix

A.1. Remaining Proofs for Submodular Sharpness

Proof of Theorem 3. Let us denote by $S_i := \{e_1, \dots, e_i\}$ the set we obtain in the i-th iteration of Algorithm 1 and $S_0 = \emptyset$. Note that $S^g := S_k$. Since the greedy algorithm chooses the element with the largest marginal in each iteration, then for all $i \in [k]$ we have

$$f(S_i) - f(S_{i-1}) \geq \max_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}(e)}$$

Now, from the submodular sharpness condition we conclude that

$$f(S_i) - f(S_{i-1}) \geq \frac{[f(S^*) - f(S_{i-1})]^{1-\theta} f(S^*)^{\theta}}{kc}.$$

The rest of the proof is the same as the proof of Theorem 1, which gives us the desired result. □

Finally, we prove the main result for the concept of dynamic submodular sharpness. This proof is similar to the proof of Theorem 2.

Proof of Theorem 4. For each iteration $i \in [k]$ in the greedy algorithm we have

$$f(S_i) - f(S_{i-1}) \geq \max_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}(e)}$$

Now, from the dynamic submodular sharpness condition we conclude that

$$f(S_i) - f(S_{i-1}) \geq \frac{[f(S^*) - f(S_{i-1})]^{1-\theta_{i-1}} f(S^*)^{\theta_{i-1}}}{kc_{i-1}},$$

which gives the same recurrence than Theorem 2. The rest of the proof is the same as the proof of Theorem 2. □

[1]Zuse Institute Berlin (ZIB), TU Berlin [2]Georgia Institute of Technology [3]CERC Data Science, Polytechnique Montreal. Correspondence to: Alfredo Torrico <alfredo.torrico-palacios@polymtl.ca>.

Proceedings of the 37^{th} International Conference on Machine Learning, Vienna, Austria, PMLR 119, 2020.

A.2. Remaining Lemmata

Lemma 1. *Consider any monotone set function $f : 2^V \to \mathbb{R}_+$. Then,*

1. *There is always a set of parameters c and θ such that f is (c, θ)-monotonic sharp. In particular, f is always (c, θ)-monotonic sharp when both $c \to 1$ and $\theta \to 0$.*
2. *If f is (c, θ)-monotonic sharp, then for any $c' \geq c$ and $\theta' \leq \theta$, f is (c', θ')-monotonic sharp. Therefore, in order to maximize the guarantee of Theorem 1 we look for the smallest feasible c and the largest feasible θ.*
3. *If f is also submodular, then Inequality* (3) *needs to be checked only for sets of size exactly k.*

Proof.

1. Note that $\frac{|S^* \setminus S|}{k} \leq 1$, so $\left(\frac{|S^* \setminus S|}{k \cdot c}\right)^{\frac{1}{\theta}} \leq \left(\frac{1}{c}\right)^{\frac{1}{\theta}}$, which shows that $\left(\frac{|S^* \setminus S|}{k \cdot c}\right)^{\frac{1}{\theta}} \to 0$ when $c \to 1$ and $\theta \to 0$. Therefore, Definition 1 is simply $\sum_{e \in S^* \setminus S} f_S(e) \geq 0$, which is satisfied since from monotonicity we have $f_S(e) \geq 0$.
2. Observe that $\left(\frac{|S^* \setminus S|}{k \cdot c}\right)^{\frac{1}{\theta}}$ as a function of c and θ is increasing in θ and decreasing in c. Therefore, $\left(\frac{|S^* \setminus S|}{k \cdot c}\right)^{\frac{1}{\theta}} \geq \left(\frac{|S^* \setminus S|}{k \cdot c'}\right)^{\frac{1}{\theta'}}$ for $c' \geq c$ and $\theta' \leq \theta$.
3. Consider a set S with i elements and such that $|S^* \setminus S| = \ell$. Let us add $k - i$ elements to S that do not belong to S^* and denote this set S'. The new set S' satisfies $|S^* \setminus S'| = \ell$, $|S'| = k$ and from submodularity we have $f_S(e) \geq f_{S'}(e)$. This proves that the inequality can be checked only for sets of size k.

□

Lemma 2. *Consider any monotone submodular set function $f : 2^V \to \mathbb{R}_+$. Then,*

1. *There is always a set of parameters c and θ such that f is (c, θ)-submodular sharp. In particular, f is always (c, θ)-submodular sharp when both $c \to 1$ and $\theta \to 0$.*

2. *If f is (c,θ)-submodular sharp, then for any $c' \geq c$ and $\theta' \leq \theta$, f is (c',θ')-submodular sharp. Therefore, in order to maximize the guarantee of Theorem 3 we look for the smallest feasible c and the largest feasible θ.*

3. *Definition 1 implies Definition 3.*

Proof. 1. Note that f satisfies the following sequence of inequalities for any set S:

$$\max_{e\in S^*\setminus S} f_S(e) \geq \frac{\sum_{e\in S^*\setminus S} f_S(e)}{|S^*\setminus S|} \geq \frac{f(S\cup S^*)-f(S)}{k} \geq \frac{f(S^*)-f(S)}{k} \tag{11}$$

where the second inequality is because of submodularity and in the last inequality we applied monotonicity. Observe that (11) is exactly (4) for $c=1$ and $\theta\to 0$.

2. Observe that $\frac{[f(S^*)-f(S)]^{1-\theta} f(S^*)^\theta}{k\cdot c}$ as a function of c and θ is increasing in θ and decreasing in c. Therefore, $\frac{[f(S^*)-f(S)]^{1-\theta} f(S^*)^\theta}{k\cdot c} \geq \frac{[f(S^*)-f(S)]^{1-\theta'} f(S^*)^{\theta'}}{k\cdot c'}$ for $c'\geq c$ and $\theta'\leq\theta$.

3. Definition 1 implies that

$$\frac{\sum_{e\in S^*\setminus S} f_S(e)}{|S^*\setminus S|} \geq \frac{\left[\sum_{e\in S^*\setminus S} f_S(e)\right]^{1-\theta} f(S^*)^\theta}{kc}. \tag{12}$$

On the other hand, by using submodularity and monotonicity we get

$$\sum_{e\in S^*\setminus S} f_S(e) \geq f(S^*\cup S)-f(S) \geq f(S^*)-f(S).$$

Therefore, by using (12) we obtain

$$\max_{e\in S^*\setminus S} f_S(e) \geq \frac{\sum_{e\in S^*\setminus S} f_S(e)}{|S^*\setminus S|} \geq \frac{[f(S^*)-f(S)]^{1-\theta} f(S^*)^\theta}{kc},$$

which proves the desired result.

□

A.3. Analysis of Monotonic Sharpness for Specific Classes of Functions

Let us denote by $\mathcal{S}(f)$ the sharpness feasible region for f, i.e., f is (c,θ)-monotonic sharp if, and only, if $(c,\theta)\in\mathcal{S}(f)$. We focus now on obtaining the best approximation guarantee for a monotone submodular function with sharpness region $\mathcal{S}(f)$.

Proposition 1. *Given a non-negative monotone submodular function $f: 2^V \to \mathbb{R}_+$ with sharpness region $\mathcal{S}(f)$, then the highest approximation guarantee $1-\left(1-\frac{\theta}{c}\right)^{\frac{1}{\theta}}$ for Problem (P_1) is given by a pair of parameters that lies on the boundary of $\mathcal{S}(f)$.*

Proof. Fix an optimal solution S^* for Problem (P_1). Note that we can compute the best pair (c,θ) for that S^* by solving the following optimization problem

$$\begin{aligned} \max \quad & 1-\left(1-\frac{\theta}{c}\right)^{\frac{1}{\theta}} \\ s.t. \quad & (c,\theta)\in\mathcal{S}(f,S^*), \end{aligned} \tag{13}$$

where $\mathcal{S}(f,S^*)$ corresponds to the sharpness region related to S^*. Observe that function $1-\left(1-\frac{\theta}{c}\right)^{\frac{1}{\theta}}$ is continuous and convex in $[1,\infty)\times(0,1]$. Note that for any $c\geq 1$, if $\theta\to 0$, then $\left(1-\frac{\theta}{c}\right)^{\frac{1}{\theta}} \to e^{-1/c}$. Also, for any subset S, Inequality (3) is equivalent to

$$\frac{|S^*\setminus S|}{k}\cdot\left(\frac{\sum_{e\in S^*\setminus S} f_S(e)}{\text{OPT}}\right)^{-\theta} - c \leq 0$$

where the left-hand side is convex as a function of c and θ, hence $\mathcal{S}(f,S^*)$ is a convex region. Therefore, the optimal pair (c^*,θ^*) of Problem (13) lies on the boundary of $\mathcal{S}(f,S^*)$. Since we considered an arbitrary optimal set, then the result easily follows. □

Let us study $\mathcal{S}(f)$ for general monotone submodular functions. If we fix $|S^*\setminus S|$, the right-hand side of (3) does not depend explicitly on S. On the other hand, for a fixed size $|S^*\setminus S|$, there is a subset S^ℓ that minimizes the left-hand side of (3), namely

$$\sum_{e\in S^*\setminus S} f_S(e) \geq \sum_{e\in S^*\setminus S^\ell} f_{S^\ell}(e),$$

for all feasible subset S such that $|S^*\setminus S| = \ell$. For each $\ell\in[k]$, let us denote

$$W(\ell) := \sum_{e\in S^*\setminus S^\ell} f_{S^\ell}(e).$$

Therefore, instead of checking Inequality (3) for all feasible subsets, we only need to check k inequalities defined by $W(1),\ldots,W(k)$. In general, computing $W(\ell)$ is difficult since we require access to S^*. However, for very small instances or specific classes of functions, this computation can be done efficiently. In the following, we provide a detailed analysis of the sharpness feasible region for specific classes of functions.

A.3.1. LINEAR FUNCTIONS.

Consider weights $w_e > 0$ for each element $e \in V$ and function $f(S) = \sum_{e \in S} w_e$. Let us order elements by weight as follows $w_1 \geq w_2 \geq \ldots \geq w_n$, where element e_i has weight w_i. We observe that an optimal set S^* for Problem (P_1) is formed by the top-k weighted elements and $\mathrm{OPT} = \sum_{i \in [k]} w_i$.

Proposition 2 (Linear functions)**.** *Consider weights $w_1 \geq w_2 \geq \ldots \geq w_n > 0$, where element $e_i \in V$ has weight w_i, and denote $W(\ell) := \sum_{i=k-\ell+1}^{k} w_i$ for each $\ell \in \{1, \ldots, k\}$. Then, the linear function $f(S) = \sum_{i : e_i \in S} w_i$ is (c, θ)-monotonic sharp in*

$$\left\{ (c,\theta) \in [1,\infty) \times [0,1] : \right. \\ \left. c \geq \frac{\ell}{k} \cdot \left(\frac{W(\ell)}{W(k)}\right)^{-\theta}, \quad \forall\, \ell \in [k-1] \right\}.$$

Moreover, this region has only $k-1$ constraints.

Proof. First, observe that $W(k) = \mathrm{OPT}$. Note that for any subset we have $|S^* \backslash S| \in \{1, \ldots, k\}$ (for $|S^* \backslash S| = 0$ the sharpness inequality is trivially satisfied). Given $\ell \in \{1, \ldots, k\}$, pick any feasible set such that $|S^* \backslash S| = \ell$, then the sharpness inequality corresponds to

$$\sum_{e \in S^* \backslash S} w_e \geq \left(\frac{\ell}{k \cdot c}\right)^{\frac{1}{\theta}} \cdot W(k), \tag{14}$$

where the left-hand side is due to linearity. Fix $\ell \in \{1, \ldots, k\}$, we observe that the lowest possible value for the left-hand side in (14) is when $S^* \backslash S = \{e_{k-\ell+1}, \ldots, e_k\}$.. Therefore, for a linear function, Definition 1 is equivalent to

$$\frac{W(\ell)}{W(k)} \geq \left(\frac{\ell}{k \cdot c}\right)^{\frac{1}{\theta}}, \quad \ell \in \{1, \ldots, k\} \quad \Leftrightarrow \\ c \geq \frac{\ell}{k} \cdot \left(\frac{W(\ell)}{W(k)}\right)^{-\theta}, \quad \ell \in \{1, \ldots, k\}.$$

Note that $\ell = k$ is redundant with $c \geq 1$. Given this, we have $k-1$ curves that define a feasible region in which the linear function is (c, θ)-monotonic sharp. In particular, if we consider $c = 1$, then we can pick $\theta = \min_{\ell \in [k-1]} \left\{ \frac{\log(k/\ell)}{\log(W(k)/W(\ell))} \right\}$. □

From Proposition 2 we observe that the sharpness of the function depends exclusively on $w_1, \ldots, w_k$. Moreover, the weights' magnitude directly affects the sharpness parameters. Let us analyze this: assume without loss of generality that $\frac{w_k}{W(k)} \leq \frac{1}{k}$, and more generally, $\frac{W(\ell)}{W(k)} \leq \frac{\ell}{k}$ for all $\ell \in [k-1]$, so we have

$$\left(\frac{\ell}{k \cdot c}\right)^{\frac{1}{\theta}} \leq \frac{W(\ell)}{W(k)} \leq \frac{\ell}{k}, \quad \ell \in \{1, \ldots, k\}$$

This shows that a sharp linear function has more similar weights in its optimal solution, i.e., when the weights in the optimal solution are *balanced*. We have the following facts for $\epsilon \in (0, 1)$:

- If $\frac{w_k}{W(k)} = (1-\epsilon) \cdot \frac{1}{k}$, then $\frac{W(\ell)}{W(k)} \geq (1-\epsilon) \cdot \frac{\ell}{k}$ for every $\ell \in [k-1]$. Observe that $c = \frac{1}{1-\epsilon}$ and $\theta = 1$ satisfy $(1-\epsilon) \cdot \frac{\ell}{k} \geq \left(\frac{\ell}{k \cdot c}\right)^{\frac{1}{\theta}}$ for any $\ell \in [k-1]$, showing that f is $\left(\frac{1}{1-\epsilon}, 1\right)$-sharp. More importantly, when ϵ is small the function becomes sharper. Also, if we set $c = 1$, then from the analysis of Proposition 2 we could pick

$$\theta = \min_{\ell \in [k-1]} \left\{ \frac{\log(k/\ell)}{\log(W(k)/W(\ell))} \right\} \\ \geq \min_{\ell \in [k-1]} \left\{ \frac{\log \frac{k}{\ell}}{\log \frac{k}{\ell(1-\epsilon)}} \right\} = \frac{\log k}{\log \frac{k}{(1-\epsilon)}},$$

showing that f is $(1, \Omega(\frac{\log k}{\log(k/(1-\epsilon))}))$-sharp. Again, when $\epsilon \to 0$ the function becomes sharper.

- On the other hand, suppose that $\frac{w_2}{W(k)} = \frac{\epsilon}{k}$, then $\frac{W(\ell)}{W(k)} \leq \epsilon \cdot \frac{\ell}{k}$ for every $\ell \in [k-1]$. Similarly to the previous bullet, by setting $c = 1$ we can choose

$$\theta = \min_{\ell \in [k-1]} \left\{ \frac{\log(k/\ell)}{\log(W(k)/W(\ell))} \right\} \\ \leq \min_{\ell \in [k-1]} \left\{ \frac{\log \frac{k}{\ell}}{\log \frac{k}{\ell\epsilon}} \right\} = \frac{\log k}{\log \frac{k}{\epsilon}},$$

showing that f is $(1, O(\frac{\log k}{\log(k/\epsilon)}))$-sharp. Observe that when $\epsilon \to 0$ the function becomes less sharp.

Observation 1. *Given parameters $c \geq 1$ and $\theta \in [0, 1]$, it is easy to construct a linear function that is (c, θ)-sharp by using Proposition 2. Without loss of generality assume $W(k) = 1$. From constraint $\ell = 1$ choose $w_k = \left(\frac{1}{kc}\right)^{\frac{1}{\theta}}$, and more generally, from constraint $\ell \in [k-1]$ choose $w_{k-\ell+1} = \left(\frac{\ell}{kc}\right)^{\frac{1}{\theta}} - \sum_{i=k-\ell+2}^{k} w_i$. Finally, set $w_1 = 1 - \sum_{i=2}^{k} w_i$.*

Observation 2. *Given $\beta \in [0, 1]$, there exists a linear function f and parameters $(c, \theta) \in [1, \infty) \times [0, 1]$ such that f is (c, θ)-sharp and $1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}} \geq 1 - \beta$. To obtain this, we use Observation 1 with $c = 1$ and any $\theta \in [0, 1]$ such that $\beta \geq (1-\theta)^{1/\theta}$.*

A.3.2. CONCAVE OVER MODULAR FUNCTIONS.

In this section, we will study a generalization of linear functions. Consider weights $w_e > 0$ for each element $e \in V$, a parameter $\alpha \in [0,1]$ and function $f(S) = \left(\sum_{e\in S} w_e\right)^\alpha$. Observe that the linear case corresponds to $\alpha = 1$. Let us order elements by weight as follows $w_1 \geq w_2 \geq \ldots \geq w_n$, where element e_i has weight w_i. Similarly than the linear case, we note that an optimal set S^* for Problem (P_1) is formed by the top-k weighted elements and $\mathrm{OPT} = \left(\sum_{i\in[k]} w_i\right)^\alpha$.

Proposition 3 (Concave over modular functions)**.** *Consider weights $w_1 \geq w_2 \geq \ldots \geq w_n > 0$, where element $e_i \in V$ has weight w_i and parameter $\alpha \in [0,1]$. Denote*

$$W(\ell) := \sum_{i=k-\ell+1}^{k} \left[\left(\sum_{j=k+1}^{k+\ell} w_j + \sum_{j=1}^{k-\ell} w_j + w_i \right)^\alpha - \left(\sum_{j=k+1}^{k+\ell} w_j + \sum_{j=1}^{k-\ell} w_j \right)^\alpha \right]$$

for each $\ell \in \{1,\ldots,k\}$. Then, the function $f(S) = \left(\sum_{i:e_i\in S} w_i\right)^\alpha$ is (c,θ)-monotonic sharp in

$$\left\{ (c,\theta) \in [1,\infty)\times[0,1] : c \geq \frac{\ell}{k}\cdot\left(\frac{W(\ell)}{\mathrm{OPT}}\right)^{-\theta}, \forall\, \ell \in [k] \right\}.$$

Proof. First, observe that unlike the linear case, $W(k) \neq \mathrm{OPT}$. Given $\ell \in \{1,\ldots,k\}$, pick any feasible set such that $|S^*\backslash S| = \ell$, then the sharpness inequality corresponds to

$$\sum_{e\in S^*\backslash S} \left(\sum_{e'\in S} w_{e'} + w_e\right)^\alpha - \left(\sum_{e'\in S} w_{e'}\right)^\alpha \geq \left(\frac{\ell}{k\cdot c}\right)^{\frac{1}{\theta}} \cdot \mathrm{OPT}\,. \quad (15)$$

Observe that function $(x+y)^\alpha - x^\alpha$ is increasing in y and decreasing in x. Therefore, the lowest possible value for the left-hand side in (15) is when $\sum_{e'\in S} w_{e'}$ is maximized and w_e is minimized. Given this, for each $\ell \in \{1,\ldots,k\}$ we choose $S = \{e_1,\ldots,e_{k-\ell},e_{k+1},\ldots,e_{k+\ell}\}$. In this way, we get $S^*\backslash S = \{e_{k-\ell+1},\ldots,e_k\}$, whose elements have the lowest weights, and S has the highest weight possible in $V\backslash\{e_{k-\ell+1},\ldots,e_k\}$. Hence, Definition 1 is equivalent to

$$\frac{W(\ell)}{\mathrm{OPT}} \geq \left(\frac{\ell}{k\cdot c}\right)^{\frac{1}{\theta}}, \quad \ell \in [k] \quad \Leftrightarrow$$
$$c \geq \frac{\ell}{k}\cdot\left(\frac{W(\ell)}{\mathrm{OPT}}\right)^{-\theta}, \quad \ell \in [k].$$

We have k curves that define a feasible region in which the function is (c,θ)-monotonic sharp with respect to S^*. In particular, if we consider $c = 1$, then we can pick $\theta = \min_{\ell\in[k]}\left\{\frac{\log(k/\ell)}{\log(\mathrm{OPT}/W(\ell))}\right\}$. □

A.3.3. COVERAGE FUNCTION, (NEMHAUSER & WOLSEY, 1978).

Consider the space $\mathcal{X} = \{1,\ldots,k\}^k$, sets $A_i = \{x \in \mathcal{X} : x_i = 1\}$ for $i \in [k-1]$ and $B_i = \{x \in \mathcal{X} : x_k = i\}$ for $i \in [k]$, ground set $V = \{A_1,\ldots,A_{k-1},B_1,\ldots,B_k\}$, and function $f(S) = \left|\bigcup_{U\in S} U\right|$ for $S \subseteq V$. In this case, Problem (P_1) corresponds to finding a family of k elements in V that covers $\mathcal{X}$ the most. By simply counting, we can see that the optimal solution for Problem (P_1) is $S^* = \{B_1,\ldots,B_k\}$ and $\mathrm{OPT} = k^k$. As shown in (Nemhauser & Wolsey, 1978), the greedy algorithm achieves the best possible $1-1/e$ guarantee for this problem.

Proposition 4. *Consider ground set $V = \{A_1,\ldots,A_{k-1},B_1,\ldots,B_k\}$. Then, the function $f(S) = \left|\bigcup_{U\in S} U\right|$ is (c,θ)-sharp in*

$$\left\{ (c,\theta) \in [1,\infty)\times[0,1] : c \geq \frac{\ell}{k}\cdot\left(\frac{\ell}{k}\cdot\left(\frac{k-1}{k}\right)^\ell\right)^{-\theta}, \quad \forall\, \ell \in [k-1] \right\}.$$

Proof. First, note that any family of the form $\{A_{i_1},\ldots,A_{i_\ell},B_{j_1},\ldots,B_{j_{k-\ell}}\}$ covers the same number of points for $\ell \in [k-1]$. Second, since there are only $k-1$ sets A_i, then any subset $S \subseteq V$ of size k satisfies $|S^*\backslash S| \leq k-1$. By simply counting, for $\ell \in [k-1]$ and set S such that $|S^*\backslash S| = \ell$, we have

$$f(S) = k^k - \ell k^{k-\ell-1}(k-1)^\ell,$$
$$f(S+e) = k^k - (\ell-1)k^{k-\ell-1}(k-1)^\ell.$$

Then, (3) becomes

$$\frac{\ell}{k}\cdot\left(\frac{k-1}{k}\right)^\ell \geq \left(\frac{\ell}{kc}\right)^{\frac{1}{\theta}}$$

Observe that f is $\left(1,\frac{1}{k}\right)$-sharp since $\ell \leq k-1$ and $\left(\frac{k-1}{k}\right)^\ell \geq \left(\frac{\ell}{k}\right)^{k-1}$. □

Observation 3 (Coverage function, (Nemhauser & Wolsey, 1978).)**.** *Note that in order to achieve $1-\left(1-\frac{\theta}{c}\right)^{\frac{1}{\theta}} \geq 1-e^{-1}$, we need $\theta \in [0,1]$ and $1 \leq c \leq \frac{\theta}{1-e^{-\theta}}$. On the other hand, by taking $\ell = k-1$ in Proposition 4 we have $c \geq \left(\frac{k-1}{k}\right)^{-k\theta+1}$, where the right-hand side tends to e^θ when k is sufficiently large. Therefore, for k sufficiently large we have $e^\theta \leq c \leq \frac{\theta}{1-e^{-\theta}}$, whose only feasible point is $c = 1$ and $\theta \to 0$.*