OPT2019: 11th Annual Workshop on Optimization for Machine Learning

On the Unreasonable Effectiveness of the Greedy Algorithm: Greedy Adapts to Sharpness

Sebastian Pokutta POKUTTA@ZIB.DE
Zuse Institute Berlin (ZIB), TU Berlin
Mohit Singh MOHIT.SINGH@ISYE.GATECH.EDU
Georgia Institute of Technology
Alfredo Torrico ALFREDO.TORRICO-PALACIOS@POLYMTL.CA
Polytechnique Montreal

Abstract

It is well known that the standard greedy algorithm guarantees a worst-case approximation factor of $1 - 1/e$ when maximizing a monotone submodular function under a cardinality constraint. However, empirical studies show that its performance is substantially better in practice. This raises a natural question of explaining this improved performance of the greedy algorithm. In this work, we define *sharpness* for submodular functions as a candidate explanation for this phenomenon. The sharpness criterion is inspired by the concept of strong convexity in convex optimization. We show that the greedy algorithm provably performs better as the sharpness of the submodular function increases. This improvement ties closely to the faster convergence rates of the first order methods for strongly convex functions. Finally, we perform a computational study to empirically support our theoretical results.

1. Introduction

During the last decade, the interest in constrained submodular maximization has increased significantly, especially due to its numerous applications in real-world problems. Common examples of these applications are influence modeling in social networks [9], sensor placement [13], document summarization [16], or in general constrained feature selection [5, 11, 14, 15, 24]. To illustrate the submodular property, consider a simple example of selecting the most influential set of nodes S in a social network where information is seeded at S and is passed around in the network based on certain stochastic process. Submodularity naturally captures the property that the total number of nodes influenced marginally decreases as more nodes are seeded [9].

The most fundamental problem in submodular optimization is to maximize a monotonically increasing submodular function subject to a cardinality constraint. It is well known that this problem is NP-hard to solve exactly. Therefore, most of the literature has focused on designing algorithms with provable guarantees. A natural approach is the standard *greedy algorithm* which constructs a set by adding in each iteration the element with the best marginal value while maintaining feasibility [21]. A classical result shows that the greedy algorithm achieves the best possible multiplicative $(1 - 1/e)$-approximation [20, 21]. However, empirical observations have shown that the greedy algorithm performs considerably better in practice. Are there specific properties in real world instances that the greedy algorithm exploits? An attempt to explain this phenomenon has been made with the concept of *curvature* [4], which measures how close to be linear is the objective function. This line of work establishes a (best possible) approximation ratio of $1 - \gamma/e$ using curvature $\gamma \in [0, 1]$ as parameter [26]. An alternative explanation can be found in the work of Chatziafratis et al. [3] who define *submodular stability*. The authors study those instances in which a unique optimal solution remains unique under multiplicative perturbations of the objective function.

In this work, we focus on giving an explanation to those instances in which the optimal solution clearly stands out over the rest of feasible solutions. For this, we consider the concept of *sharpness* initially presented in continuous optimization [17] and we adapt it to submodular optimization. Sharpness in continuous optimization translates in faster convergence rates. Equivalently, we will show that in submodular maximization the greedy algorithm performs better as the sharpness of the objective function increases. Our main contributions in this work are: (1) to introduce the *sharpness* criteria in submodular optimization as a novel candidate explanation of the performance of the greedy algorithm; (2) to show that the standard greedy algorithm automatically adapt to the sharpness of the objective function, without requiring this information as part of the input; (3) to provide provable guarantees that depend on the sharpness parameters for this algorithm; and (4) to empirically support our theoretical results with real-world applications.

1.1. Problem Formulation

In this work, we study the *submodular function maximization problem subject to a single cardinality constraint*. Formally, consider a ground set of n elements $V = \{1, \dots, n\}$ and a non-negative set function $f : 2^V \to \mathbb{R}_+$. We denote the marginal value for any subset $A \subseteq V$ and $e \in V$ by $f_A(e) := f(A + e) - f(A)$, where $A + e := A \cup \{e\}$. A set function f is *submodular* if, and only if, it satisfies the *diminishing returns property*. Namely, for any $e \in V$ and $A \subseteq B \subseteq V \backslash \{e\}$, $f_A(e) \geq f_B(e)$. We say that f is *monotone* if for any $A \subseteq B \subseteq V$, we have $f(A) \leq f(B)$. For simplicity, let us assume that $f(\emptyset) = 0$. Given $k \in \mathbb{Z}_+$ and a non-negative monotone submodular function f, we focus on the following optimization problem:

$$\max\{f(S) : |S| \leq k\}. \tag{P$_1$}$$

We denote the optimal value as OPT. In this context, we assume the *value oracle model*, i.e., the decision-maker queries the value of S and the oracle returns $f(S)$. As we mentioned earlier, to explain the improved performance of the greedy algorithm, we borrow the concept of *sharpness* used in continuous optimization [1, 7, 17, 18, 23]. Broadly speaking, this property characterizes the behavior of a function around the set of optimal solutions. Sharpness has been widely used to study convergence rates in convex and non-convex optimization, see e.g. [2, 8, 10, 22, 25]. For a detailed review on the sharpness condition in continuous optimization, we refer the interested reader to [25].

1.2. Our Contributions and Results

Our main contribution is to introduce the concept of submodular sharpness and show that the greedy algorithm performs better than the worst-case guarantee $1 - 1/e$ for functions that are sharp with appropriate parameters.

Given parameters $c \geq 1$ and $\theta \in [0, 1]$, we define submodular sharpness as follows.

Definition 1 (Submodular Sharpness) *A non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ is said to be (c, θ)-sharp, if there exists an optimal solution S^* for Problem (P$_1$) such that for any subset $S \subseteq V$ with $|S| \leq k$ the function satisfies*

$$\sum_{e \in S^* \backslash S} f_S(e) \geq \left(\frac{|S^* \backslash S|}{k \cdot c} \right)^{\frac{1}{\theta}} \cdot \mathrm{OPT}\,. \tag{1}$$

The property can be interpreted as implying that the optimal set S^* is not just unique but any solution which differs significantly from S^* has substantially lower value. Our first main result for Problem (P_1) is stated in the next theorem.

Theorem 2 *Consider a non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ which is (c, θ)-sharp. Then, the greedy algorithm returns a feasible set S^g for (P_1) such that*

$$f(S^g) \geq \left[1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}}\right] \cdot f(S^*).$$

We remark that *any* monotone submodular function f is (c, θ)-sharp as $c \leftarrow 1$ and $\theta \leftarrow 0$ (Lemma 5 in the Appendix). Thus, Theorem 2 implies that the greedy algorithm gives a $1 - \frac{1}{e}$-approximation for any monotone submodular function recovering the classical guarantee [21]. But if the parameters are bounded away from $(1, 0)$, we obtain a strict improvement over this worst case guarantee. In Section 3, we show experimental results to illustrate that real data sets do show improved parameters.

1.2.1. DYNAMIC SHARPNESS

Definition 1 can be considered as a *static* notion of sharpness, since parameters c and θ do not change with respect to the size of S. We generalize this definition by considering the notion of *dynamic sharpness*, in which the parameters c and θ depend on the size of the feasible sets, i.e., $c_{|S|} \geq 1$ and $\theta_{|S|} \in [0, 1]$.

Definition 3 (Dynamic Submodular Sharpness) *A non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ is said to be* dynamic (c, θ)-sharp, *where $c = (c_0, c_1, \ldots, c_{k-1}) \in [1, \infty)^k$ and $\theta = (\theta_0, \theta_1, \ldots, \theta_{k-1}) \in [0, 1]^k$, if there exists an optimal solution S^* for Problem (P_1) such that for any subset $S \subseteq V$ with $|S| \leq k$ the function satisfies*

$$\sum_{e \in S^* \backslash S} f_S(e) \geq \left(\frac{|S^* \backslash S|}{k \cdot c_{|S|}}\right)^{\frac{1}{\theta_{|S|}}} \cdot f(S^*)$$

In other words, we say that a function f is (c_i, θ_i)-sharp for any subset such that $|S| = i$, where $i \in \{0, \ldots, k-1\}$. Note that since we have k pairs of parameters (c_i, θ_i), then there are at most $k - 1$ intervals in which sharpness may change. If the parameters are identical in each interval, then we recover Definition 1 of submodular sharpness. We obtain the following guarantee for dynamic sharpness.

Theorem 4 *Consider a non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ that is dynamic (c, θ)-sharp with parameters $c = (c_0, c_1, \ldots, c_{k-1}) \in [1, \infty)^k$ and $\theta = (\theta_0, \theta_1, \ldots, \theta_{k-1}) \in [0, 1]^k$. Then, the greedy algorithm returns a set S^g for (P_1) such that*

$$f(S^g) \geq \left[1 - \left(\left(\left(1 - \frac{\theta_0}{c_0 k}\right)^{\frac{\theta_1}{\theta_0}} - \frac{\theta_1}{c_1 k}\right)^{\frac{\theta_2}{\theta_1}} - \cdots - \frac{\theta_{k-1}}{c_{k-1} k}\right)^{\frac{1}{\theta_{k-1}}}\right] \cdot f(S^*).$$

Due to space constraints, we provide the proofs of Theorem 2 and Theorem 4 in the Appendix.

2. Contrasting Sharpness with Curvature

A natural question is how our results compare to the curvature analysis proposed in [4]. Specifically, is there any class of functions in which the sharpness criterion provides considerable better guarantees than the curvature analysis? The total curvature of a monotone submodular function $f : 2^V \to \mathbb{R}_+$ is defined as $\gamma = 1 - \min_{e \in V^*}\{f_{V-e}(e)/f(e)\}$, where $V^* = \{e \in V : f(e) > 0\}$. Conforti and Cornuéjols [4] proved that if a monotone submodular function has curvature $\gamma \in [0, 1]$, then the standard greedy algorithm guarantees a $(1 - e^{-\gamma})/\gamma$ fraction of the optimal value. Let us contrast the two concepts. Consider an integer $k \geq 2$, a ground set $V = [n]$ with $n = 2k$, a monotone submodular set function $f(S) = \min\{|S|, k + 1\}$ and problem (P_1). Observe that any set of size k is an optimal set, so consider S^* any set with k elements. Note also that the curvature of the function is $\gamma = 1$, since $f(V) = k + 1$ and $f(V - e) = k + 1$ for any $e \in V$. Therefore, the curvature analysis guarantees a $1 - 1/e$ factor. Let us analyze the sharpness of this function. Pick any subset $S \subseteq V$ such that $|S| \leq k$ and $S \neq S^*$, then we have $f(S) = |S|$ and $f_S(e) = 1$ for any $e \in S^* \backslash S$. Hence,

$$\frac{\sum_{e \in S^* \backslash S} f_S(e)}{f(S^*)} = \frac{|S^* \backslash S|}{k},$$

which implies that parameters $\theta = 1$ and $c = 1$ are feasible in the sharpness inequality (1). Therefore, the sharpness analysis gives us an approximation factor of 1. From this simple example, we observe that curvature is a global parameter of the function that does not take into consideration the optimal set and can be easily perturbed, while the sharpness criterion focuses on the behavior of the function around the optimal solution. More precisely, take any function f with curvature close to 0, which means an approximation guarantee close to 1. Then, take $\tilde{f}(S) = \min\{f(S), f(S^*)\}$. This function is still monotone and submodular, but its curvature now is 1, while its sharpness is the same as the original function f.

3. Experimental Results

In this section, we provide a computational study of the sharpness criterion in real-world applications. We aim to explicitly obtain the sharpness parameters of the objective function for different small ground sets. With these results, we will empirically show how the approximation factors vary with respect to the cardinality budget k. We contrast these results with the approximation ratios guaranteed by the curvature analysis [4] and submodular stability [3]. We also analyze the concept of dynamic sharpness: either with 1 or $k - 1$ intervals. We will observe that the approximation factors are considerable improved compared to the static sharpness results. We provide extra computational studies in the Appendix.

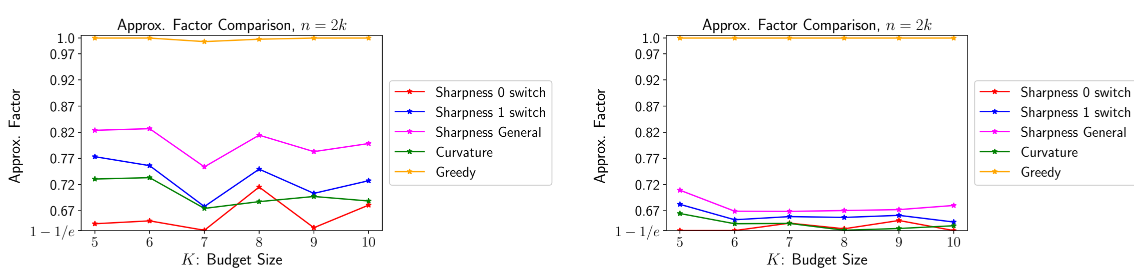
3.1. Non-parametric Learning

For this application we follow the setup in [19]. Let X_V be a set of random variables corresponding to bio-medical measurements, indexed by a ground set of patients V. We assume X_V to be a Gaussian Process (GP), i.e., for every subset $S \subseteq V$, X_S is distributed according to a multivariate normal distribution $\mathcal{N}(\mu_S, \mathbf{\Sigma}_{S,S})$, where $\mu_S = (\mu_e)_{e \in S}$ and $\mathbf{\Sigma}_{S,S} = [\mathcal{K}_{e,e'}]_{e,e' \in S}$ are the prior mean vector and prior covariance matrix, respectively. The covariance matrix is given in terms of a positive definite kernel $\mathcal{K}$, e.g., a common choice in practice is the squared exponential kernel $\mathcal{K}_{e,e'} = \exp(-\|x_e - x_{e'}\|_2^2/h)$. Most efficient approaches for making predictions in GPs rely on choosing a small subset of data points. For instance, in the Informative Vector Machine (IVM) the goal is to obtain a subset A such that maximizes the information gain, $f(A) = \frac{1}{2} \log \det(\mathbf{I} + \sigma^{-2}\, \mathbf{\Sigma}_{A,A})$, which is monotone and submodular [12]. In our experiment, we use the Parkinson Telemonitoring dataset [27] consisting of a total of $5,875$ patients with early-stage Parkinson's disease and the corresponding bio-medical voice measurements with 22 attributes (dimension of the observations). We normalize the vectors to zero mean and unit norm. With these measurements we computed the covariance matrix Σ considering the squared exponential kernel with parameter $h = 0.75$. For the objective function we consider $\sigma = 1$. As we mentioned before, the objective in this application is to select the k most informative patients. The objective function in this experiment is highly non-linear which makes difficult to obtain the sharpness parameters. Therefore, for this experiment we consider different small random instances with $n = 2k$ where $k = \{5, \dots, 10\}$. In Figure 1 (a), we plot the variation of the approximation factors with respect to different values of k. The best approximation factor is obtained when using the general dynamic sharpness approach, which clearly outperforms the curvature analysis. Observe that the best approximation factor obtained by considering one moment in which the sharpness change (blue) also improves the curvature analysis. We also plot the ratios obtained by the greedy solution and we observe that most of the time it is close to 1.

3.2. Movie Recommendation

For this application we consider the MovieLens data-set [6] which consists of 7,000 users and 13,977 movies. Each user had to rank at least one movie with an integer value in $\{0, \dots, 5\}$ where 0 denotes that the movies was not ranked by that user. Therefore, we have a matrix $[r_{ij}]$ of rankings for each user i and each movie j. The objective in this application is to select the k highest ranked movies among the users. To make the computations less costly in terms of time, we use only $m = 1000$ users. In the same spirit, we will choose a small number n from the 13,977 movies.

In this section, we consider the facility-location function $f(S) = \frac{1}{m}\sum_{i\in[m]} \max_{j\in S} r_{ij}$. This function is known to be non-negative, monotone, and submodular. Most of the time this function is not submodular stable [3] since it has multiple optimal solutions. It is not clear how to efficiently obtain the sharpness parameters of this function, therefore for this experiment, we consider different small random instances with $n = 2k$ where $k = \{5, \dots, 10\}$. In Figure 1 (b), we plot the variation of the approximation factors with respect to different values of k. We observed that the greedy algorithm (orange) always finds an optimal solution. We note that the curvature analysis does not provide a good approximation ratio. Instead, we obtain a considerable improvement if we use the general dynamic sharpness approach.



(a) (b)

Figure 1: Variations in the approximation factors with respect to different budgets k: *Non-parametric learning* (a) and *Facility-location function* (b).

References

[1] Jérôme Bolte, Aris Daniilidis, and Adrian Lewis. The Lojasiewicz inequality for nonsmooth subanalytic functions with applications to subgradient dynamical systems. *SIAM Journal on Optimization*, 17(4):1205–1223, 2007.

[2] Jérôme Bolte, Shoham Sabach, and Marc Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1):459–494, 2014.

[3] Vaggos Chatziafratis, Tim Roughgarden, and Jan Vondrák. Stability and recovery for independence systems. In *Proceedings of the 25th European Symposium on Algorithms (ESA)*, 2017.

[4] Michele Conforti and Gérard Cornuéjols. Submodular set functions, matroids and the greedy algorithm: Tight worst-case bounds and some generalizations of the rado-edmonds theorem. *Discrete Applied Mathematics*, 7(3):251 – 274, 1984.

[5] Abhimanyu Das and David Kempe. Algorithms for subset selection in linear regression. In *Proceedings of the 40th Annual ACM Symposium on the Theory of Computing (STOC)*, pages 45–54, 2008.

[6] F. Maxwell Harper and Joseph A. Konstan. The movielens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4):19:1–19, 2016.

[7] Alan J. Hoffman. On approximate solutions of systems of linear inequalities. *Mathematical Programming*, 49(4):263–265, 1952.

[8] Hamed Karimi, Julie Nutini, and Mark Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition. In *Machine Learning and Knowledge Discovery in Databases*, pages 795–811, 2016.

[9] David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2003.

[10] Thomas Kerdreux, Alexandre d'Aspremont, and Sebastian Pokutta. Restarting Frank-Wolfe. arXiv, 2018.

[11] Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 324–331, 2005.

[12] Andreas Krause and Carlos Guestrin. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the Twenty-First Conference on Uncertainty in Artificial Intelligence*, UAI'05, pages 324–331. AUAI Press, 2005.

[13] Andreas Krause, H. Brendan McMahan, Carlos Guestrin, and Anupam Gupta. Robust submodular observation selection. *Journal of Machine Learning Research*, 9:2761–2801, 2008.

[14] Andreas Krause, Ajit Singh, and Carlos Guestrin. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008.

[15] Andreas Krause, Ram Rajagopal, Anupam Gupta, and Carlos Guestrin. Simultaneous placement and scheduling of sensors. In *Proceedings of the 8th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN)*, pages 181–192, 2009.

[16] Hui Lin and Jeff A. Bilmes. How to select a good training-data subset for transcription: submodular active selection for sequences. In *Proceedings of the 10th Annual Conference of the International Speech Communication Association (INTERSPEECH)*, pages 2859–2862, 2009.

[17] Stanislas Lojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, pages 1575–1595, 1963.

[18] Stanislas Lojasiewicz. Sur la géométrie semi-et sous-analytique. *Annales de l'Institut Fourier*, 43(5):1575–1595, 1993.

[19] Baharan Mirzasoleiman, Ashwinkumar Badanidiyuru, Amin Karbasi, Jan Vondrák, and Andreas Krause. Lazier than lazy greedy. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, pages 1812–1818, 2015.

[20] George L. Nemhauser and Laurence A. Wolsey. Best algorithms for approximating the maximum of a submodular set function. *Mathematics of Operations Research*, 3(3):177–188, 1978.

[21] George L. Nemhauser, Laurence A. Wolsey, and Marshall L. Fisher. An analysis of approximations for maximizing submodular set functions–I. *Mathematical Programming*, 14(1):265–294, 1978.

[22] Arkadi S. Nemirovskii and Yurii E. Nesterov. Optimal methods of smooth convex minimization. *USSR Computational Mathematics and Mathematical Physics*, 25(2):21–30, 1985.

[23] B. Polyak. Sharp minima. Presented in Workshop on generalized Lagrangians and their applications (IIASA), 1979.

[24] Thomas Powers, Jeff Bilmes, David W. Krout, and Les Atlas. Constrained robust submodular sensor selection with applications to multistatic sonar arrays. In *Proceedings of the 19th International Conference on Information Fusion (FUSION)*, pages 2179–2185, 2016.

[25] Vincent Roulet and Alexandre d'Aspremont. Sharpness, restart and acceleration. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, pages 1119–1129, 2017.

[26] Maxim Sviridenko, Jan Vondrák, and Justin Ward. Optimal approximation for submodular and supermodular optimization with bounded curvature. In *Proceedings of the 26th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1134–1148, 2015.

[27] A. Tsanas, M. A. Little, P. E. McSharry, and L. O. Ramig. Enhanced classical dysphonia measures and sparse regression for telemonitoring of parkinson's disease progression. In *Proceedings of the 35th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 594–597, 2010.

Appendix A. Appendix

A.1. Remaining Proofs

Lemma 5 *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)-sharp. In particular, f is always (c, θ)-sharp when both $c \to 1$ and $\theta \to 0$.*

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

3. *Inequality (1) needs to be checked only for sets of size exactly k.*

Proof

1. Note that $\frac{|S^*\backslash S|}{k} \leq 1$, so $\left(\frac{|S^*\backslash S|}{k \cdot c}\right)^{\frac{1}{\theta}} \leq \left(\frac{1}{c}\right)^{\frac{1}{\theta}}$, which shows that $\left(\frac{|S^*\backslash S|}{k \cdot c}\right)^{\frac{1}{\theta}} \to 0$ when $c \to 1$ and $\theta \to 0$. Therefore, Definition 1 is simply $\sum_{e \in S^*\backslash S} f_S(e) \geq 0$, which is satisfied since from monotonicity we have $f_S(e) \geq 0$.

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

3. Given that f is submodular, then $f_A(e) \geq f_B(e)$ for any $e \in V$ and $A \subseteq B \subseteq V \backslash e$, so in (1) we can assume without loss of generality that $|S| = k$.

■

We are ready to prove Theorem 2. We emphasize that the greedy algorithm automatically adapts to the sharpness of the function and does not require explicit access to the sharpness parameters in order to obtain the desired guarantees. For completeness, let us recall the standard greedy algorithm.

Algorithm 1: Greedy [21]

Input: ground set $V = \{1, \ldots, n\}$, monotone submodular function $f : 2^V \to \mathbb{R}_+$, and $k \in \mathbb{Z}_+$.
Output: feasible set S with $|S| \leq k$.

Initialize $S = 0$.
while $|S| < k$ **do**
    $S \leftarrow S + \text{argmax}_{e \in V \backslash S} f_S(e)$
end while

Recall that given parameters $c \geq 1$ and $\theta \in [0,1]$, a function is (c,θ)-sharp if there exists an optimal set S^* such that for any set S with at most k elements, then

$$\sum_{e \in S^* \setminus S} f_S(e) \geq \left(\frac{|S^* \setminus S|}{k \cdot c} \right)^{\frac{1}{\theta}} \cdot f(S^*)$$

Proof [Proof of Theorem 2] Let us denote by S_i the set we obtain in the i-th iteration of Algorithm 1. Note that $S^g := S_k$. By using the properties of the function f, we can obtain the following sequence of inequalities

$$\begin{aligned} f(S_i) - f(S_{i-1}) &= \frac{\sum_{e \in S^* \setminus S_{i-1}} f(S_i) - f(S_{i-1})}{|S^* \setminus S_{i-1}|} && (2) \\ &\geq \frac{\sum_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}}(e)}{|S^* \setminus S_{i-1}|} && \text{(choice of greedy)} \end{aligned}$$

Now, from the sharpness condition we know that

$$\frac{1}{|S^* \setminus S_{i-1}|} \geq \frac{f(S^*)^\theta}{kc} \cdot \left(\sum_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}}(e) \right)^{-\theta}$$

so we obtain the following bound

$$\begin{aligned} f(S_i) - f(S_{i-1}) &\geq \frac{f(S^*)^\theta}{kc} \cdot \left(\sum_{e \in S^* \setminus S_{i-1}} f_{S_{i-1}}(e) \right)^{1-\theta} && \text{(sharpness)} \\ &\geq \frac{f(S^*)^\theta}{kc} \cdot [f(S_{i-1} \cup S^*) - f(S_{i-1})]^{1-\theta} && \text{(submodularity)} \\ &\geq \frac{f(S^*)^\theta}{kc} \cdot [f(S^*) - f(S_{i-1})]^{1-\theta}. && \text{(monotonicity)} \end{aligned}$$

Therefore, we need to solve the following recurrence

$$a_i \geq a_{i-1} + \frac{a^\theta}{kc} \cdot [a - a_{i-1}]^{1-\theta} \qquad (3)$$

where $a_i = f(S_i)$, $a_0 = 0$ and $a = f(S^*)$.

Define $h(x) = x + \frac{a^\theta}{kc} \cdot [a-x]^{1-\theta}$, where $x \in [0,a]$. Observe that $h'(x) = 1 - \frac{a^\theta(1-\theta)}{kc} \cdot [a-x]^{-\theta}$. Therefore, h is increasing in the interval $I := \left\{ x : \ 0 \leq x \leq a \cdot \left(1 - \left(\frac{1-\theta}{kc}\right)^{1/\theta} \right) \right\}$. Let us define

$$b_i := a \cdot \left[1 - \left(1 - \frac{\theta}{c} \cdot \frac{i}{k} \right)^{\frac{1}{\theta}} \right].$$

First, let us check that $b_i \in I$ for all $i \in \{0, \ldots, k\}$. Namely, for any i we need to show that

$$a \cdot \left[1 - \left(1 - \frac{\theta}{c} \cdot \frac{i}{k} \right)^{\frac{1}{\theta}} \right] \leq a \cdot \left(1 - \left(\frac{1-\theta}{kc} \right)^{1/\theta} \right) \quad \Leftrightarrow \quad (kc - i\theta) \geq 1 - \theta$$

The expression $kc - i\theta$ is decreasing on i. Hence, we just need the inequality for $i = k$, namely $k(c-\theta) \geq 1-\theta$, which is true since $c \geq 1$ and $k \geq 1$.

Our goal is to prove that $a_i \geq b_i$, so by induction let us assume that $a_{i-1} \geq b_{i-1}$ is true (observe $a_0 \geq b_0$). By using monotonicity of h on the interval I, we get $h(a_{i-1}) \geq h(b_{i-1})$. Also, observe that recurrence (3) is equivalent to write $a_i \geq h(a_{i-1})$ which implies that $a_i \geq h(b_{i-1})$. To finish the proof we will show that $h(b_{i-1}) \geq b_i$.

Assume for simplicity that $a = 1$. For $x \in [0, k]$, define

$$g(x) := \left(1 - \frac{\theta}{kc} \cdot x\right)^{1/\theta}.$$

Note that $g'(x) = -\frac{1}{kc} \cdot g(x)^{1-\theta}$ and $g''(x) = \frac{1-\theta}{(kc)^2} \cdot g(x)^{1-2\theta}$. Observe that g is convex, so for any $x_1, x_2 \in [0, k]$ we have $g(x_2) \geq g(x_1) + g'(x_1) \cdot (x_2 - x_1)$. By considering $x_2 = i$ and $x_1 = i - 1$, we obtain

$$g(i) - g(i-1) - g'(i-1) \geq 0 \tag{4}$$

On the other hand,

$$\begin{aligned} h(b_{i-1}) - b_i &= 1 - \left(1 - \frac{\theta}{c} \cdot \frac{i-1}{k}\right)^{\frac{1}{\theta}} + \frac{1}{kc} \cdot \left(1 - \frac{\theta}{c} \cdot \frac{i-1}{k}\right)^{\frac{1-\theta}{\theta}} - 1 + \left(1 - \frac{\theta}{c} \cdot \frac{i}{k}\right)^{\frac{1}{\theta}} \\ &= \left(1 - \frac{\theta}{c} \cdot \frac{i}{k}\right)^{\frac{1}{\theta}} - \left(1 - \frac{\theta}{c} \cdot \frac{i-1}{k}\right)^{\frac{1}{\theta}} + \frac{1}{kc} \cdot \left(1 - \frac{\theta}{c} \cdot \frac{i-1}{k}\right)^{\frac{1-\theta}{\theta}} \end{aligned}$$

which is exactly the left-hand side of (4), proving $h(b_{i-1}) - b_i \geq 0$.

Finally,

$$f(S^g) = a_k \geq b_k = \left[1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}}\right] \cdot f(S^*),$$

proving the desired guarantee. ■

We recover the classical $1 - 1/e$ approximation factor, originally proved in [21].

Corollary 6 *The greedy algorithm achieves a $1 - \frac{1}{e}$-approximation for any monotone submodular function.*

Proof We prove in Lemma 5 (see Appendix) that any monotone submodular function is (c, θ)-sharp when $\theta \to 0$ and $c \to 1$. On the other hand, we know that $1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}}$ is increasing on θ, so for all $c \geq 1$ and $\theta \in [0, 1]$ we have $1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}} \geq 1 - e^{-1/c}$. By taking limits, we obtain $\lim_{c \to 1, \theta \to 0} 1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}} \geq 1 - e^{-1}$. ■

Let us denote by $\mathcal{S}(f)$ the sharpness feasible region for f, i.e., f is (c, θ)-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 7 *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

$$\begin{aligned} \max \quad & 1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}} \\ s.t. \quad & (c, \theta) \in \mathcal{S}(f, S^*), \end{aligned} \tag{5}$$

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 (1) is equivalent to

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

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 (5) 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 $|S^* \backslash S|$, the right-hand side of (1) does not depend explicitly on S. On the other hand, for a fixed size $|S^* \backslash S|$, there is a subset S^ℓ that minimizes the left-hand side of (1), namely

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

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

$$W(\ell) := \sum_{e \in S^* \backslash S^\ell} f_{S^\ell}(e). \tag{6}$$

Therefore, instead of checking Inequality (1) 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.

A.1.1. DYNAMIC SHARPNESS

In this section, we focus on proving the main results for dynamic sharpness, Theorem 4. We emphasize that the greedy algorithm automatically adapts to the dynamic sharpness of the function without requiring parameters (c_i, θ_i) as part of the input.

Proof [Proof of Theorem 4] Observe that in the i-th iteration of the greedy algorithm $|S_i| = i$, so the change of the sharpness parameters will occur in every iteration i. The proof is similar to Theorem 2, but the recursion needs to be separated in each step i. Let us recall recursion (3): for any $i \in [k]$

$$a_i \geq a_{i-1} + \frac{a^\theta}{k c_{i-1}} \cdot [a - a_{i-1}]^{1-\theta_{i-1}},$$

where $a_i = f(S_i)$, $a_0 = 0$, and $a = f(S^*)$. For simplicity assume that $a = 1$. We proceed the proof by induction. Note that for $i = 1$ we need to prove that $a_1 \geq \frac{1}{kc_0}$. For c_0 and θ_0, the sharpness inequality (1) needs to be checked only for $S = \emptyset$, which is trivially satisfied with $c_0 = \theta_0 = 1$. From the proof of Theorem 2 we can conclude the following for $i = 1$.

$$a_1 \geq \left[1 - \left(1 - \frac{\theta_0}{kc_0}\right)^{\frac{1}{\theta_0}}\right],$$

and given that $c_0 = \theta_0 = 1$ is valid pair of parameters, then this inequality is simply $a_1 \geq \frac{1}{kc_0}$, proving the desired base case. Let us denote

$$b_j := \left[1 - \left(\left(\left(1 - \frac{\theta_0}{c_0 k}\right)^{\frac{\theta_1}{\theta_0}} \cdots - \frac{\theta_{j-2}}{c_{j-2}k}\right)^{\theta_{j-1}/\theta_{j-2}} - \frac{\theta_{j-1}}{c_{j-1}k}\right)^{\frac{1}{\theta_{j-1}}}\right]$$

for $1 \leq j \leq k$. We assume that $a_i \geq b_i$ is true, and will prove that $a_{i+1} \geq b_{i+1}$. Similarly to the proof of Theorem 2, we define $h(x) := x + \frac{1}{kc_i}[1-x]^{1-\theta_i}$ for $x \in [0,1]$, which is increasing in the interval $I := \left\{x :\ 0 \leq x \leq 1 - \left(\frac{1-\theta_i}{kc_i}\right)^{1/\theta_i}\right\}$. Let us prove that $b_i \in I$. First, observe that $b_i \geq 0$. For the other inequality in I we have

$$b_i \leq 1 - \left(\frac{1-\theta_i}{kc_i}\right)^{1/\theta_i} \quad \Leftrightarrow \quad \left(\left(1 - \frac{\theta_0}{c_0 k}\right)^{\theta_1/\theta_0} \cdots - \frac{\theta_{i-1}}{c_{i-1}k}\right)^{\theta_i/\theta_{i-1}} \geq \frac{1-\theta_i}{kc_i},$$

which is satisfied for sufficiently large k.

Similarly than the proof of Theorem 2, for $x \in [i, k]$ define

$$g(x) := \left(\left(\left(1 - \frac{\theta_0}{c_0 k}\right)^{\frac{\theta_1}{\theta_0}} \cdots - \frac{\theta_{i-1}}{c_{i-1}k}\right)^{\frac{\theta_i}{\theta_{i-1}}} - \frac{\theta_i}{c_i k} \cdot (x - i)\right)^{\frac{1}{\theta_i}}$$

Note that $g'(x) = -\frac{1}{kc_i} \cdot g(x)^{1-\theta_i}$ and $g''(x) = \frac{1-\theta_i}{(kc_i)^2} \cdot g(x)^{1-2\theta_i}$. Observe that g is convex, so for any $x_1, x_2 \in [i, k]$ we have $g(x_2) \geq g(x_1) + g'(x_1) \cdot (x_2 - x_1)$. By considering $x_2 = i + 1$ and $x_1 = i$, we obtain

$$g(i+1) - g(i) - g'(i) \geq 0 \tag{7}$$

Inequality (7) is exactly $h(b_i) - b_{i+1} \geq 0$, since $g(i+1) = 1 - b_{i+1}$ and $g(i) = 1 - b_i$. Finally, since we assumed $a_i \geq b_i$, then $a_{i+1} \geq h(a_i) \geq h(b_i) \geq b_{i+1}$, where the first inequality is the definition of the recursion, the second inequality is due to the monotonicity of h in the interval I, and finally, the last inequality was proven in (7). Therefore, $a_k \geq b_k$ which proves the desired guarantee since $f(S^g) = a_k$.

■

Observation 1 *Note that we recover Theorem 2 when $(c_i, \theta_i) = (c, \theta)$ for all $i \in \{0, \ldots, k-1\}$.*

A.2. Computational Study

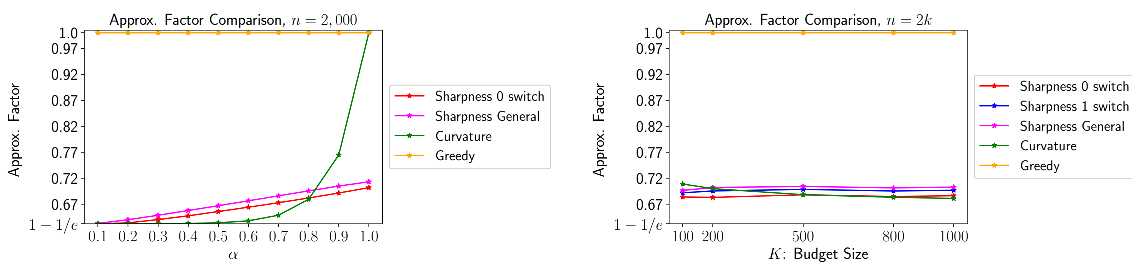
Before providing the algorithm to find the sharpness of a function we observe the following: in Inequality (2) of the proof of Theorem 2, we divide by $|S^*\backslash S_{i-1}|$ which is valid for iterations $i \in \{1, \ldots, k\}$. Specifically, in iteration $i = k$, we will divide by $|S^*\backslash S_{k-1}|$, which later will be replaced by the sharpness criterion. Note that $|S_{k-1}| = k - 1$, which implies that numerically it is sufficient enough to have the sharpness criterion for subsets $|S| \leq k - 1$. Moreover, equivalently to Fact 3. in Lemma 5, we actually need to check sets of size exactly $k - 1$. In a similar way, when we face two set of parameters (c, θ) and (c', θ'), we apply the same logic for the switching moment $t \in [k]$: we check only subsets of size exactly $t - 1$.

Simple search for sharpness parameters. Fix an optimal solution S^*. Compute $W(\ell)$ for all $\ell \in [k]$ as described in (6). To find parameters (c, θ) we follow a simple search over the boundary of $\mathcal{S}(f)$ defined by S^*. For c, we sequentially iterate over possible values in a fixed range $[1, c_{max}]$ (we consider granularity 0.01 and $c_{max} = 3$). Given c, we compute $\theta = \min_{\ell\in[k]} \left\{\frac{\log(kc/\ell)}{\log(\mathrm{OPT}/W(\ell))}\right\}$. Once we have c and θ, we compute the corresponding approximation factor. If we improve we continue and update c; otherwise, we stop. A similar procedure is done for the case of dynamic sharpness.

A.2.1. MOVIE RECOMMENDATION: EXTRA RESULTS

We follow the same setup as in Section 3.2. In our first experiment, we consider the following function $f(S) = \left(\frac{1}{m}\sum_{i\in[m]}\sum_{j\in S} r_{ij}\right)^\alpha$ where $\alpha \in (0, 1]$. We consider a random instance of $n = 2,000$ movies and we want to select $k = 1,000$ movies. We take 10 different values for $\alpha \in \{0.1, 0.2, \ldots, 1.0\}$. First, we noticed in our experiment that the instance is not submodular stable [3] since it had multiple optimal solutions. In Figure 2 (a) we plot the variation of the approximation factors with respect to different values of α. We observe that there is an α^* in which the curvature analysis improves the sharpness criterion. Below this α^*, the dynamic sharpness with $k - 1$ switching points (magenta) clearly outperforms the curvature analysis; similarly, for the static sharpness. We conclude that the curvature analysis is ideal when the objective function is closer to be a linear function.

In the next experiment, we fix $\alpha = 0.8$ and we consider different random instances with $n = 2k$, where $k = \{50, 100, 200, 500, 800, 1000\}$. In Figure 2 (b) we plot the variation of the approximation factors with respect to different k's. We observe that the general dynamic sharpness outperforms the curvature analysis for k sufficiently large. We also plot in blue the best approximation factor that can be obtained by using a single switching moment. The results for this approach are significantly close to the general dynamic sharpness.



(a) (b)

Figure 2: *Concave over modular function:* (a) variations in the approximation factors with respect to different exponents α, (b) $\alpha = 0.8$, variations in the approximation factors with respect to different budgets k.