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

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

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. We show that the greedy algorithm provably performs better as the sharpness of the submodular function increases. This improvement ties in closely with the faster convergence rates of first order methods for sharp functions in convex optimization.

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 (Kempe et al., 2003), sensor placement (Krause et al., 2008a), document summarization (Lin & Bilmes, 2009), or in general constrained feature selection (Krause & Guestrin, 2005a; Das & Kempe, 2008; Krause et al., 2008b; 2009; Powers et al., 2016). To illustrate the submodular property, consider a simple example of selecting the most influential nodes S in a social network where information is seeded at S and is passed around in the network based on a certain stochastic process. Submodularity captures the natural property that the total number of nodes influenced marginally decreases as more nodes are seeded (Kempe et al., 2003). Given the importance of submodular optimization, there has been significant progress in designing new algorithms with provable guarantees (Calinescu et al., 2011; Ene & Nguyen, 2016; Buchbinder & Feldman, 2016; Sviridenko, 2004).

The most fundamental problem in submodular optimization is to maximize a monotonically increasing submodular function subject to a cardinality constraint. A classical result (Nemhauser et al., 1978; Nemhauser & Wolsey, 1978) shows that the greedy algorithm is a multiplicative $(1-1/e)$-approximation algorithm. Moreover, no other efficient algorithm can obtain a better guarantee (Nemhauser & Wolsey, 1978). However, empirical observations have shown that standard algorithms such as the greedy algorithm performs considerably better in practice. Explaining this phenomenon has been a tantalizing challenge. 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* (Conforti & Cornuéjols, 1984). In simple words, this parameter measures how close to linear the objective function is. This line of work establishes a (best possible) approximation ratio of $1-\gamma/e$ using curvature $\gamma \in [0,1]$ as parameter (Sviridenko et al., 2015).

In this work, we focus on giving an explanation for 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 introduced in continuous optimization (Łojasiewicz, 1963) and we adapt it to submodular optimization. Roughly speaking, this property measures the behavior of the objective function around the set of optimal solutions. Sharpness in continuous optimization translates in faster convergence rates. Equivalently, we will show that the greedy algorithm for submodular maximization performs better as the sharpness of the objective function increases, as a discrete analog of ascent algorithms in continuous optimization.

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 adapts to the sharpness of the objective func-

[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, Online, PMLR 119, 2020. Copyright 2020 by the author(s).

tion, without requiring this information as part of the input; (3) to provide provable guarantees that depend only on the sharpness parameters; and (4) to empirically support our theoretical results with a detailed computational study in 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 $V = \{1, \ldots, n\}$ 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 $f(A) \leq f(B)$. To ease the notation, we will write the value of singletons as $f(e) := f(\{e\})$. For simplicity, we assume throughout this work that f is *normalized*, i.e., $f(\emptyset) = 0$. Our results still hold when $f(\emptyset) \neq 0$, but an additive extra term has to be carried over.

As we mentioned before, our work is mostly focused on the optimization of non-negative monotone submodular functions subject to a single cardinality constraint. In this setting, we are given a non-negative integer k and the goal is to optimally select a subset S that contains at most k elements of V. Formally, the optimization problem is the following

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

Throughout the rest of the paper, 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 $f(S)$. It is well known that (P_1) is NP-hard to solve exactly under the value oracle model. 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 (Fisher et al., 1978). The authors show that the greedy algorithm achieves a $1 - 1/e$ approximation factor for problem (P_1), which is tight (Nemhauser & Wolsey, 1978; Feige, 1998). We give a detailed description of the related work in Section 1.3. Even though the best possible guarantee is $1 - 1/e$, the standard greedy algorithm usually performs better in practice.

To explain this phenomenon we adapt the concept of *sharpness* used in continuous optimization (Hoffman, 1952; Łojasiewicz, 1963; Polyak, 1979; Lojasiewicz, 1993; Bolte et al., 2007). The notion of sharpness is also known as Hölderian error bound on the distance to the set of optimal solutions. Broadly speaking, this property characterizes the behavior of a function around the set of optimal solutions. The sharpness criteria along with other similar conditions have been proposed in the continuous optimization literature for which better convergence rates are obtained (Karimi et al., 2016). In this work, we will adapt some of these notions to their discrete analog in submodular optimization. To exemplify these properties, consider a concave function F, a feasible region X, a set of optimal solutions $X^* = \mathrm{argmax}_{x \in X} F(x)$ and a distance function $d(\cdot, X^*) : X \to \mathbb{R}_{\geq 0}$. Then, F is said to be (c, θ)-sharp if for any $x \in X$

$$F(x^*) - F(x) \geq \left(\frac{d(x, X^*)}{c}\right)^{1/\theta},$$

where $F^* = \max_{x \in X} F(x)$. Since F is concave, we we have

$$\nabla F(x) \cdot (x^* - x) \geq \left(\frac{d(x, X^*)}{c}\right)^{1/\theta}. \tag{1}$$

Sharpness implies another related condition: the Polyak-Łojasiewicz (PL) inequality (Polyak, 1963; Łojasiewicz, 1963) or Gradient Dominated property. We say that the function F satisfies the PL inequality if there exists $c \geq 1$ and $\theta \in [0, 1]$ such that for all $x \in X$

$$\|\nabla F(x)\| \geq \frac{1}{c}(F(x^*) - F(x))^{1-\theta}. \tag{2}$$

As we mentioned earlier sharpness and similar criteria have been widely used to study convergence rates in convex and non-convex optimization, see e.g., (Nemirovskii & Nesterov, 1985; Karimi et al., 2016; Bolte et al., 2014; Roulet & d'Aspremont, 2017; Kerdreux et al., 2019). For a detailed review on the sharpness condition in continuous optimization, we refer the interested reader to (Roulet & d'Aspremont, 2017).

1.2. Our Contributions and Results

Our main contribution is to introduce multiple concepts of sharpness in submodular optimization that mimic conditions previously studied in continuous optimization. We show that the greedy algorithm performs better than the worst-case guarantee $1 - 1/e$ for functions that are sharp with appropriate parameters. Empirically, we obtain improved guarantees on real data sets when using more refined conditions.

1.2.1. Monotonic Sharpness

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

Definition 1 (Monotonic Sharpness)**.** *A non-negative monotone submodular set function $f : 2^V \to \mathbb{R}_+$ is said to be* (c, θ)-monotonic 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*

$$\sum_{e \in S^* \backslash S} f_S(e) \geq \left(\frac{|S^* \backslash S|}{k \cdot c} \right)^{\frac{1}{\theta}} \cdot \mathrm{OPT}\,. \tag{3}$$

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. Observe that Inequality (3) resembles Inequality (1), where the marginal values around the optimal solution play the role of the gradient of the function and the distance to the optimal set corresponds to the cardinality of the difference. Our first main result for Problem (P_1) is stated in the next theorem.

Theorem 1. *Consider a non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ which is (c, θ)-monotonic 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 set function f is (c, θ)-monotonic sharp for a pair of parameters satisfying $c \to 1$ and $\theta \to 0$ (Lemma 1 in the Appendix). Corollary 1 in Section 2 shows that the guarantee $1 - (1 - \theta/c)^{1/\theta}$ tends to $1 - 1/e$ when (c, θ) goes to $(1, 0)$, recovering the classical guarantee for any monotone submodular function (Nemhauser et al., 1978). However, if the parameters (c, θ) 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. In the Appendix, we also discuss the sharpness of simple submodular functions such as linear and concave over linear functions.

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 monotonic sharpness*, in which the parameters c and θ depend on the size of the feasible sets, i.e., $c_{|S|} \geq 1$ and $\theta_{|S|} \in [0, 1]$. This allows us to obtain improved guarantees for the greedy algorithm based on how the monotonic sharpness changes dynamically. Formally, we define dynamic sharpness as follows.

Definition 2 (Dynamic Monotonic Sharpness)**.** *A non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ is said to be* dynamic (c, θ)-monotonic 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)-monotonic 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 every interval, then we recover Definition 1 of monotonic sharpness. We obtain the following guarantee for dynamic monotonic sharpness.

Theorem 2. *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*

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

In Section 3, we empirically show that the guarantees shown in Theorem 2 strictly outperforms the factors provided by Theorem 1.

1.2.2. Submodular Sharpness

Definition 1 measures the distance between a feasible set S and S^* as the cardinality of its difference. However, this distance may not be a precise measure, since the value $f(S)$ could be quite close to OPT. Therefore, we introduce the concept of *submodular sharpness* as a natural generalization of Definition 1. To ease the notation, in this section we will use the same letters c and θ for the parameters which not necessarily correspond to the parameters in Definition 1. Given parameters $c \geq 1$ and $\theta \in [0, 1]$, we define submodular sharpness as follows

Definition 3 (Submodular Sharpness)**.** *A non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ is said to be* (c, θ)-submodular 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*

$$\max_{e \in S^* \backslash S} f_S(e) \geq \frac{1}{kc} \left[f(S^*) - f(S) \right]^{1-\theta} \mathrm{OPT}^{\theta} \tag{4}$$

Inequality (4) can be interpreted as the submodular version of the Polyak-Łojasiewicz inequality (2) with respect to the ℓ_∞-norm, where the marginal values play the role of the gradient of the function and we consider a general exponent θ. We observe that any non-negative monotone submodular function that is (c, θ)-monotonic sharp is also (c, θ)-submodular sharp (see Lemma 2 in the Appendix). We obtain the following main result for submodular sharpness.

Theorem 3. *Consider a non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ which is (c, θ)-submodular sharp. Then, the greedy algorithm returns a feasible set S^g for (P_1) such that*

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

Observe that *any* monotone submodular function f is (c, θ)-submodular sharp as $c \leftarrow 1$ and $\theta \leftarrow 0$ (see Lemma 2 in the Appendix). Thus, Theorem 3 recovers the classical guarantee $1 - 1/e$ (Nemhauser et al., 1978). The reader might think that the approximation guarantee in Theorem 3 is the same as in Theorem 1, but this is not the case. Since Definition 3 is weaker than Definition 1 (see Lemma 2 in the Appendix), the approximation factor guaranteed by Theorem 3 is at least as good as in Theorem 1. More importantly, we will empirically show in Section 3 that there is a significant improvement in the approximation guarantees when using Definition 3.

Similar to dynamic monotonic sharpness, we introduce the concept of *dynamic submodular sharpness*.

Definition 4 (Dynamic Submodular Sharpness)**.** *A non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ is said to be* dynamic (c, θ)-submodular 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*

$$\max_{e \in S^* \setminus S} f_S(e) \geq \frac{1}{kc_{|S|}} \left[f(S^*) - f(S)\right]^{1-\theta_{|S|}} f(S^*)^{\theta_{|S|}} \tag{5}$$

Finally, we obtain the following result for Problem (P_1).

Theorem 4. *Consider a non-negative monotone submodular function $f : 2^V \to \mathbb{R}_+$ that is dynamic (c, θ)-submodular 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 prove Theorems 3 and 4 in the Appendix.

Experimental Results. In Section 3, we provide a computational study in real-world applications such as movie recommendation, non-parametric learning, and clustering. We emphasize that our goal is to experimentally verify the performance of the sharpness guarantees and contrast our theoretical results with the existing literature, such as the concepts of *curvature* (Conforti & Cornuéjols, 1984) and *submodular stability* (Chatziafratis et al., 2017). While all these results try to explain the improved performance of greedy, sharpness provides an alternate explanation for this improved behavior. Given that the sharpness definitions require an optimal solution, experiments on large datasets are not possible. However, we stress that the greedy algorithm does not require prior knowledge of the sharpness parameters. In addition, we show in Appendix A.3 that the computation of monotonic sharpness can be done efficiently in some specific classes of monotone submodular functions.

1.3. Related Work

As remarked earlier, the greedy algorithm gives a $(1 - \frac{1}{e})$-approximation for maximizing a submodular function subject to a cardinality constraint (Nemhauser et al., 1978) and is optimal (Feige, 1998; Nemhauser & Wolsey, 1978).

The concept of *curvature*, introduced in (Conforti & Cornuéjols, 1984), measures how far the function is from being linear. Formally, a monotone submodular function $f : 2^V \to \mathbb{R}_+$ has *total curvature* $\gamma \in [0, 1]$ if

$$\gamma = 1 - \min_{e \in V^*} \frac{f_{V-e}(e)}{f(e)}, \tag{6}$$

where $V^* = \{e \in V : \ f(e) > 0\}$. For a submodular function with total curvature $\gamma \in [0, 1]$, Conforti and Cornuéjols (1984) showed that the greedy algorithm guarantees an approximation factor of $(1 - e^{-\gamma})/\gamma$. The notion of curvature has been also used when minimizing submodular functions (Iyer et al., 2013), and the equivalent notion of *steepness* in supermodular function minimization (Il'ev, 2001); as well as maximizing submodular functions under general combinatorial constraints (Conforti & Cornuéjols, 1984; Sviridenko, 2004; Sviridenko et al., 2015; Feldman, 2018); we refer to the interested reader to the literature therein for more details.

Close to our setting is the concept of stability widely studied in discrete optimization. Broadly speaking, there are instances in which the unique optimal solution still remains unique even if the objective function is slightly perturbed. For example, the concept of *clusterability* has been widely studied in order to show the existence of *easy instances* in clustering (Balcan et al., 2009; Daniely et al., 2012), we refer the interested reader to the survey (Ben-David, 2015) for more details. Stability has been also studied in other contexts such as influence maximization (He & Kempe, 2014), Nash equilibria (Balcan & Braverman, 2017), and Max-Cut (Bilu & Linial, 2012). Building on (Bilu & Linial, 2012), the concept of stability under multiplicative perturbations in submodular optimization is studied in (Chatziafratis et al., 2017). Formally, given a non-negative monotone submodular function f, $\tilde{f}$ is a γ-perturbation if: (1) $\tilde{f}$ is non-negative monotone submodular; (2) $f \leq \tilde{f} \leq \gamma \cdot f$; and (3) for any $S \subseteq V$ and $e \in V \backslash S$, $0 \leq \tilde{f}_S(e) - f_S(e) \leq (\gamma - 1) \cdot f(e)$. Now, assume we have an instance of problem (P_1) with a unique optimal solution, then this instance is said to be γ-stable if for any γ-perturbation of the objective function, the original optimal solution remains being unique. Chatziafratis et al. (2017) show that the greedy algorithm recovers the unique optimal solution for 2-stable instances. However, it is not hard to show that 2-stability for problem (P_1) is a strong assumption, since 2-stable instances can be easily solved by maximizing the sum of singleton values and thus 2-stable functions do not capture higher order relationship among elements.

2. Monotonic Sharpness Analysis

In this section, we focus on the analysis of the standard greedy algorithm for Problem (P_1) when the objective function is (c, θ)-monotonic sharp. 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, we recall the standard greedy algorithm in Algorithm 1.

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

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

Algorithm 1 Greedy (Nemhauser et al., 1978)

Input: ground set $V = \{1, \ldots, n\}$, monotone submodular function $f : 2^V \to \mathbb{R}_+$, and $k \in \mathbb{Z}_+$.
Output: set S with $|S| \leq k$.
1: Initialize $S = 0$.
2: **while** $|S| < k$ **do**
3: $\quad S \leftarrow S + \operatorname{argmax}_{e \in V \backslash S} f_S(e)$
4: **end while**

Proof of Theorem 1. 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

$$f(S_i) - f(S_{i-1}) = \frac{\sum_{e \in S^* \backslash S_{i-1}} f(S_i) - f(S_{i-1})}{|S^* \backslash S_{i-1}|} \quad (7)$$

$$\geq \frac{\sum_{e \in S^* \backslash S_{i-1}} f_{S_{i-1}}(e)}{|S^* \backslash S_{i-1}|} \quad \text{(choice of greedy)}$$

Now, from the sharpness condition we know that

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

so we obtain the following bound

$$f(S_i) - f(S_{i-1}) \geq \frac{f(S^*)^\theta}{kc} \cdot \left(\sum_{e \in S^* \backslash S_{i-1}} f_{S_{i-1}}(e) \right)^{1-\theta} \quad \text{(sharpness)}$$

$$\geq \frac{f(S^*)^\theta}{kc} \cdot [f(S_{i-1} \cup S^*) - f(S_{i-1})]^{1-\theta} \quad \text{(submodularity)}$$

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

Therefore, we need to solve the following recurrence

$$a_i \geq a_{i-1} + \frac{a^\theta}{kc} \cdot [a - a_{i-1}]^{1-\theta} \quad (8)$$

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

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

$$(kc - i\theta) \geq 1 - \theta$$

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 by induction that $a_i \geq b_i$. First, let us prove that $a_1 \geq b_1$. Since $a_0 = 0$, the recursion implies that

$a_1 \geq a/kc$. On the other hand, note that $b_1 = a \cdot [1 - (1 - \theta/kc)^{1/\theta}]$. Given that $(1 - \theta/kc)^{1/\theta}$ is decreasing in θ, we conclude that $a_1 \geq b_1$.

Now, let us assume that $a_{i-1} \geq b_{i-1}$ is true and prove that $a_i \geq b_i$. For this part, we consider the case in which $a_{i-1} \in I$ for all $i \in \{1, \ldots, k\}$. If this is not the case, namely, there exists $i \in [k]$ such that $a_{i-1} \notin I$, then we have that

$$a_{i-1} > a \cdot \left(1 - \left(\frac{1-\theta}{kc}\right)^{1/\theta}\right),$$

and since $a_k \geq a_{i-1}$ (because of monotonicity of the function f) we obtain a $1 - \left(\frac{1-\theta}{kc}\right)^{1/\theta}$ approximation factor, which is better than the guarantee we want to prove. Therefore, in the worst-case $a_{i-1} \in I$ for all $i \in \{1, \ldots, k\}$. Given the monotonicity of h on the interval I and the inductive hypothesis, we get $h(a_{i-1}) \geq h(b_{i-1})$. Also, observe that recurrence (8) 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

$$g(i) - g(i-1) - g'(i-1) \geq 0 \tag{9}$$

On the other hand,

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

which is exactly the left-hand side of (9), 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. □

As we mentioned earlier, we recover the classical $1 - 1/e$ approximation factor, originally proved in (Nemhauser et al., 1978).

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

Proof. We prove in Lemma 1 (see Appendix) that any monotone submodular function is (c, θ)-monotonic 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 $\lim_{c \to 1, \theta \to 0} 1 - \left(1 - \frac{\theta}{c}\right)^{\frac{1}{\theta}} \geq 1 - e^{-1}$. □

2.1. Monotonic Dynamic Sharpness

In this section, we focus on proving the main results for dynamic sharpness, Theorem 2. 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 of Theorem 2. 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 1, but the recursion needs to be separated in each step i. Let us recall recursion (8): for any $i \in [k]$

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

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 (3) needs to be checked only for $S = \emptyset$, which is trivially satisfied with $c_0 = \theta_0 = 1$. From the proof of Theorem 1, 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 1, 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

$$b_i \leq 1 - \left(\frac{1-\theta_i}{kc_i}\right)^{1/\theta_i} \quad \Leftrightarrow$$

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

which is satisfied for sufficiently large k.

Similarly than the proof of Theorem 1, 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

$$g(i+1) - g(i) - g'(i) \geq 0 \tag{10}$$

Inequality (10) 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 (10). Therefore, $a_k \geq b_k$ which proves the desired guarantee since $f(S^g) = a_k$. □

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

2.2. Contrasting Sharpness with Curvature

A natural question is how our results compare to the curvature analysis proposed in (Conforti & Cornuéjols, 1984). Specifically, is there any class of functions in which the sharpness criterion provides considerable better guarantees than the curvature analysis? 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 (3). 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 non-negative monotone submodular 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. Conforti and Cornuéjols (1984) also define the notion of *greedy curvature* in which the parameter is computed only with respect to the subsets constructed by the greedy algorithm. This notion does provide better guarantees than the total curvature and our previous example does not apply. Similarly, we could define *greedy sharpness*, but we do not pursue this since the definition would become algorithmically dependent. One of the main contributions of this work is that the greedy algorithm automatically adapts to sharpness without prior knowledge or dependency between these two.

3. Computational Study

In this section, we provide a computational study of the sharpness criteria in three real-world applications: movie recommendation, non-parametric learning and exemplar-based clustering. In these experiments, 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 different instances defined by the cardinality budget k. We will observe that the curvature analysis (Conforti & Cornuéjols, 1984), submodular stability (Chatziafratis et al., 2017) and monotonic sharpness are not enough, but more refined concepts as dynamic monotonic sharpness and submodular sharpness in its two versions provide strictly better results.

Search for monotonic sharpness. Fix an optimal solution S^*. For each $\ell \in [k]$, compute

$$W(\ell) := \min_S \left\{ \sum_{e \in S^* \backslash S} f_S(e) : |S| \leq k, \ |S^* \backslash S| = \ell \right\}.$$

To find parameters (c, θ) we follow a simple search: we sequentially iterate over possible values of c 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.

Search for submodular sharpness. Fix an optimal solution S^*. To find parameters (c, θ) we follow a simple search: we sequentially iterate over possible values of c in a fixed range $[1, c_{max}]$ (we consider granularity 0.01 and $c_{max} = 3$). Given c, we compute $\theta = \min_{|S|\le k}\left\{\frac{\log(kcW_2(S)/W(S))}{\log(\mathrm{OPT}/W(S))}\right\}$, where

$$W(S) := \mathrm{OPT} - f(S) \quad \text{and} \quad W_2(S) := \max_{e\in S^*\setminus S} f_S(e).$$

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 submodular sharpness.

Experiments setup. For each application, we will run experiments on small ground sets. For each budget size k, we sample $n = 2k$ elements from the data sets which will be considered as the ground set. In each graph, we will plot the approximation factor (y-axis) obtained by the corresponding method in each instance (x-axis). The analysis we will study are: curvature (defined in Section 1.3), monotonic and submodular sharpness (computed as described above), and finally, the greedy ratio (worst possible value output by the greedy algorithm in the corresponding instance).

3.1. Non-parametric Learning

For this application we follow the setup in (Mirzasoleiman et al., 2015). 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 was shown to be monotone and submodular in (Krause & Guestrin, 2005b).

In our experiment, we use the Parkinson Telemonitoring dataset (Tsanas et al., 2010) 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 nonlinear 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 instances of size $n = 2k$. Observe that the greedy algorithm finds a nearly optimal solution in each instance. The best approximation factor is obtained when using the concept of dynamic submodular sharpness, which is considerably close to the greedy results. These results significantly improve the ones obtained by the curvature analysis and monotonic sharpness, providing evidence that more refined notions of sharpness can capture the behavior of the greedy algorithm.

3.2. Movie Recommendation

For this application we consider the MovieLens data-set (Harper & Konstan, 2016) 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 $m = 1000$ random users. In the same spirit, we will choose a small number n from the 13,977 movies.

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 $\alpha = 0.8$ and different small random instances with $n = 2k$ where $k = \{5, \dots, 10\}$. First, we noticed in our experiment that the instance is not submodular stable (Chatziafratis et al., 2017) since it had multiple optimal solutions. In Figure 1 (b) we plot the variation of the approximation factors with respect to different k's. We observe that monotonic sharpness already gives us improved guarantees with respect to the worst-case $1 - 1/e$, although worse results than the curvature analysis. More refined definitions as the submodular sharpness slightly improve the results for any instance.

In the next experiment, 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 2-stable (Chatziafratis et al., 2017) since it has multiple optimal solutions. For this function, we consider different small random instances with $n = 2k$ elements in the ground set where $k = \{5, \dots, 10\}$. In Figure 1 (c) 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 and monotonic sharpness barely improve the worst-case ratio $1 - 1/e$. We

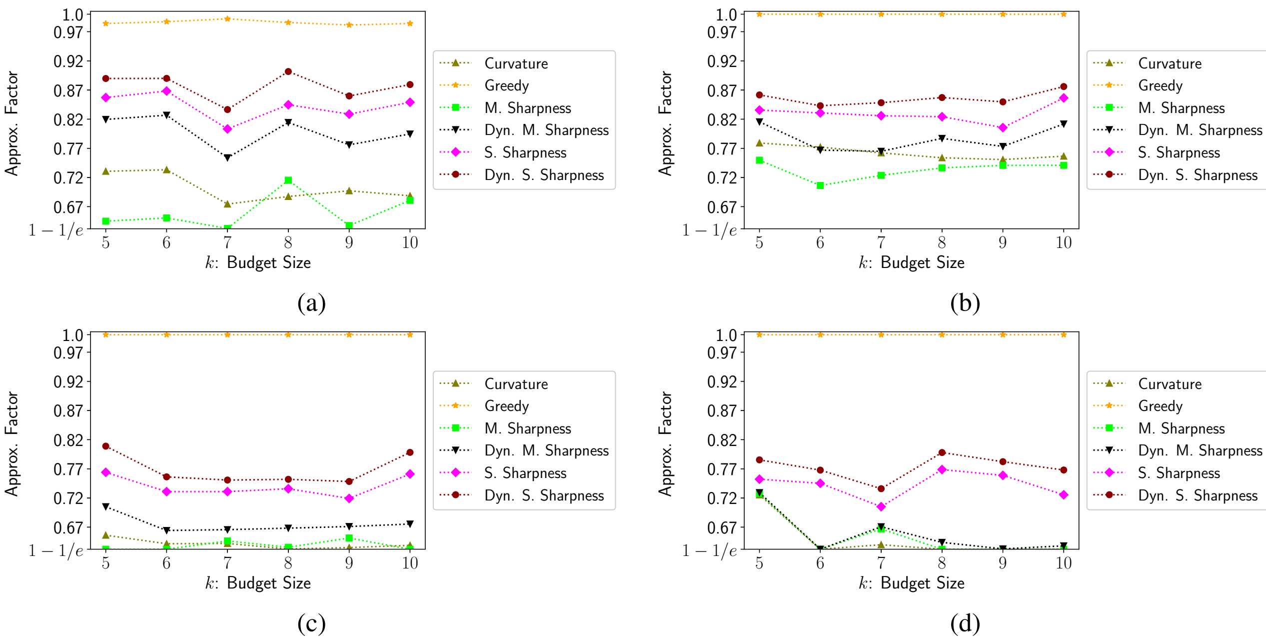

Figure 1. Approximation factors with respect to different budgets k. *Non-parametric learning:* (a); *Movie recommendation:* (b) concave over modular with $\alpha = 0.8$ and (c) facility location function; *Exemplar-based clustering:* (d).

obtain a significant improvement if we use the submodular sharpness approach. However, the gap between the greedy results and the dynamic submodular sharpness is still substantial, which may be due to the shape of this objective function: facility-location functions are known to be *flat* and have multiple optimal solutions.

3.3. Exemplar-based Clustering

We follow the setup in (Mirzasoleiman et al., 2015). Solving the k-medoid problem is a common way to select a subset of exemplars that represent a large dataset V (Kaufman & Rousseeuw, 2009). This is done by minimizing the sum of pairwise dissimilarities between elements in $A \subseteq V$ and V. Formally, define $L(A) = \frac{1}{V}\sum_{e\in V}\min_{v\in A} d(e, v)$, where $d : V \times V \to \mathbb{R}_+$ is a distance function that represents the dissimilarity between a pair of elements. By introducing an appropriate auxiliary element e_0, it is possible to define a new objective $f(A) := L(\{e_0\}) - L(A + e_0)$ that is monotone and submodular (Gomes & Krause, 2010), thus maximizing f is equivalent to minimizing L.

In our experiment, we use the VOC2012 dataset (Everingham et al., 2012) which contains around 10,000 images. The ground set V corresponds to images, and we want to select a subset of the images that best represents the dataset. Each image has several (possible repeated) associated categories such as person, plane, etc. There are around 20 categories in total. Therefore, images are represented by feature vectors obtained by counting the number of elements that belong to each category, for example, if an image has 2 people and one plane, then its feature vector is $(2, 1, 0, \dots, 0)$ (where zeros correspond to other elements). We choose the Euclidean distance $d(e, e') = \|x_e - x_{e'}\|$ where $x_e, x_{e'}$ are the feature vectors for images e, e'. We normalize the feature vectors to mean zero and unit norm, and we choose e_0 as the origin.

For this experiment, we consider different random small instances with $n = 2k$ where $k = \{5, \dots, 10\}$. The objective function in this experiment is non-linear which makes difficult to obtain the sharpness parameters. In Figure 1 (d) we plot the variation of the approximation factors with respect to different instances defined by k. The dynamic submodular sharpness approach outperforms the rest of the procedures, although the greedy algorithm always finds an optimal solution.

Acknowledgements

We thank the reviewers and the meta-reviewer for the time spent reviewing our manuscript and their helpful comments. Research reported in this paper was partially supported by NSF grant CCF-BSF-AF: 1717947, CCF-AF:1910423 and NSF CAREER award CMMI-1452463.

References

Balcan, M.-F. and Braverman, M. Nash equilibria in pertubation-stable games. *Theory of Computing*, 13(13):

1–31, 2017.

Balcan, M.-F., Blum, A., and Gupta, A. Approximate clustering without the approximation. In *Proceedings of the 20th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 1068–1077, 2009.

Ben-David, S. Computational feasibility of clustering under clusterability assumptions. *CoRR*, abs/1501.00437, 2015.

Bilu, Y. and Linial, N. Are stable instances easy? *Combinatorics, Probability and Computing*, 21(5):643–660, 2012.

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

Bolte, J., Sabach, S., and Teboulle, M. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146(1):459–494, 2014.

Buchbinder, N. and Feldman, M. Deterministic algorithms for submodular maximization problems. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pp. 392–403, 2016.

Calinescu, G., Chekuri, C., Pál, M., and Vondrák, J. Maximizing a monotone submodular function subject to a matroid constraint. *SIAM Journal on Computing*, 40(6): 1740–1766, 2011.

Chatziafratis, V., Roughgarden, T., and Vondrák, J. Stability and recovery for independence systems. In *Proceedings of the 25th European Symposium on Algorithms (ESA)*, 2017.

Conforti, M. and Cornuéjols, G. 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.

Daniely, A., Linial, N., and Saks, M. E. Clustering is difficult only when it does not matter. *CoRR*, abs/1205.4891, 2012.

Das, A. and Kempe, D. Algorithms for subset selection in linear regression. In *Proceedings of the 40th Annual ACM Symposium on the Theory of Computing (STOC)*, pp. 45–54, 2008.

Ene, A. and Nguyen, H. L. Constrained submodular maximization: Beyond 1/e. In *Proceedings of the 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pp. 248–257, 2016.

Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results. http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html, 2012.

Feige, U. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4):634–652, 1998.

Feldman, M. Guess free maximization of submodular and linear sums. arXiv, 2018.

Fisher, M. L., Nemhauser, G. L., and Wolsey, L. A. An analysis of approximations for maximizing submodular set functions–II. In *Polyhedral combinatorics*, pp. 73–87. Springer, 1978.

Gomes, R. and Krause, A. Budgeted nonparametric learning from data streams. In *Proceedings of the 27th International Conference on Machine Learning (ICML)*, pp. 391–398, 2010.

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

He, X. and Kempe, D. Stability of influence maximization. In *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, pp. 1256–1265, 2014.

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

Il'ev, V. P. An approximation guarantee of the greedy descent algorithm for minimizing a supermodular set function. *Discrete Applied Mathematics*, 114(1):131 – 146, 2001.

Iyer, R. K., Jegelka, S., and Bilmes, J. A. Curvature and optimal algorithms for learning and minimizing submodular functions. In *Proceedings of the 27th International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 2742–2750, 2013.

Karimi, H., Nutini, J., and Schmidt, M. Linear convergence of gradient and proximal-gradient methods under the Polyak-Lojasiewicz condition. In *Machine Learning and Knowledge Discovery in Databases*, pp. 795–811, 2016.

Kaufman, L. and Rousseeuw, P. J. *Finding groups in data: an introduction to cluster analysis*, volume 344. Wiley-Interscience, 2009.

Kempe, D., Kleinberg, J., and Tardos, E. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2003.

Kerdreux, T., d'Aspremont, A., and Pokutta, S. Restarting frank-wolfe. In *Proceedings of the 22nd International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 89, pp. 1275–1283, 2019.

Krause, A. and Guestrin, C. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 324–331, 2005a.

Krause, A. and Guestrin, C. Near-optimal nonmyopic value of information in graphical models. In *Proceedings of the 21st Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 324–331, 2005b.

Krause, A., McMahan, H. B., Guestrin, C., and Gupta, A. Robust submodular observation selection. *Journal of Machine Learning Research*, 9:2761–2801, 2008a.

Krause, A., Singh, A., and Guestrin, C. Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies. *Journal of Machine Learning Research*, 9:235–284, 2008b.

Krause, A., Rajagopal, R., Gupta, A., and Guestrin, C. Simultaneous placement and scheduling of sensors. In *Proceedings of the 8th ACM/IEEE Conference on Information Processing in Sensor Networks (IPSN)*, pp. 181–192, 2009.

Lin, H. and Bilmes, J. A. 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)*, pp. 2859–2862, 2009.

Łojasiewicz, S. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, pp. 1575–1595, 1963.

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

Mirzasoleiman, B., Badanidiyuru, A., Karbasi, A., Vondrák, J., and Krause, A. Lazier than lazy greedy. In *Proceedings of the 29th AAAI Conference on Artificial Intelligence (AAAI)*, pp. 1812–1818, 2015.

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

Nemhauser, G. L., Wolsey, L. A., and Fisher, M. L. An analysis of approximations for maximizing submodular set functions–I. *Mathematical Programming*, 14(1):265–294, 1978.

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

Polyak, B. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3:864–878, 12 1963.

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

Powers, T., Bilmes, J., Krout, D. W., and Atlas, L. Constrained robust submodular sensor selection with applications to multistatic sonar arrays. In *Proceedings of the 19th International Conference on Information Fusion (FUSION)*, pp. 2179–2185, 2016.

Roulet, V. and d'Aspremont, A. Sharpness, restart and acceleration. In *Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS)*, pp. 1119–1129, 2017.

Sviridenko, M. A note on maximizing a submodular set function subject to a knapsack constraint. *Operations Research Letters*, 32(1):41–43, 2004.

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

Tsanas, A., Little, M. A., McSharry, P. E., and Ramig, L. O. 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)*, pp. 594–597, 2010.