Submodular Norms with Applications To Online Facility Location and Stochastic Probing

Kalen Patton
School of Mathematics, Georgia Tech, Atlanta, GA, USA

Matteo Russo
DIAG, Sapienza Università di Roma, Italy

Sahil Singla
School of Computer Science, Georgia Tech, Atlanta, GA, USA

Abstract

Optimization problems often involve vector norms, which has led to extensive research on developing algorithms that can handle objectives beyond ℓ_p norms. Our work introduces the concept of *submodular norms*, which are a versatile type of norms that possess marginal properties similar to submodular set functions. We show that submodular norms can either accurately represent or approximate well-known classes of norms, such as ℓ_p norms, ordered norms, and symmetric norms. Furthermore, we establish that submodular norms can be applied to optimization problems such as online facility location and stochastic probing. This allows us to develop a logarithmic-competitive algorithm for online facility location with symmetric norms, and to prove logarithmic adaptivity gap for stochastic probing with symmetric norms.

2012 ACM Subject Classification Theory of computation → Online algorithms

Keywords and phrases Submodularity, Monotone Norms, Online Facility Location, Stochastic Probing

Digital Object Identifier 10.4230/LIPIcs.APPROX/RANDOM.2023.23

Category APPROX

Funding *Matteo Russo*: Supported by the ERC Advanced Grant 788893 AMDROMA.

Acknowledgements We thank the anonymous reviewers of APPROX/RANDOM 2023 for their valuable feedback.

1 Introduction

In the field of combinatorial optimization, norm objectives are frequently encountered. Canonical problems, such as the min-weight spanning tree and the k-median, involve searching for a feasible solution that minimizes the sum of costs, which is equivalent to the ℓ_1 norm of the edge cost vector. On the other hand, canonical problems like bottleneck spanning tree and k-center aim to minimize the maximum of costs, which is equivalent to the ℓ_∞ norm of the edge cost vector. However, because ℓ_1 and ℓ_∞ norms only capture the extreme Utilitarian and Egalitarian objectives respectively, significant research has been devoted to developing combinatorial optimization algorithms for more general norms (see references in Section 1.3). Among the commonly studied norms are ℓ_p norms, ordered norms, Orlicz norms, symmetric norms, and arbitrary monotone norms.

Over the past decade, there is also a lot of effort towards designing online and stochastic algorithms for more general norms. For instance, remarkable progress has been made in developing algorithms beyond ℓ_p norms for various problems, such as load balancing [16, 17, 29, 30, 31, 32], set cover [7, 36], spanning trees [28], and bandits with knapsacks [33, 32]. Notably, most of the recent progress is for the class of symmetric norms, i.e.,

© Kalen Patton, Matteo Russo, and Sahil Singla;
licensed under Creative Commons License CC-BY 4.0
Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2023).
Editors: Nicole Megow and Adam D. Smith; Article No. 23; pp. 23:1–23:22
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

monotone norms that remain unchanged upon permutation of coordinates. This progress is partly due to Ky Fan's Dominance Theorem (refer to [11]), which reduces the problem of designing algorithms for symmetric norms to ordered norms (see Section 1.1 for a formal definition). Ordered norms are comparatively more manageable due to their explicit form. Given this progress on some combinatorial problems for symmetric norms, a natural question arises:

What general norms and what combinatorial problems admit algorithms with good performance guarantees?

A challenge in making progress beyond symmetric norms is that such norms are not explicit, e.g., they may not be well approximated by ordered norms. In this work we introduce the class of *submodular norms*, which is a broad class of norms with marginals properties mimicking submodular set functions. We show that submodular norms either capture or approximate popular classes of norms like ℓ_p norms, ordered norms, and symmetric norms. Moreover, submodular norms are amenable to some of the optimization problems that were previously intractable like online facility location and stochastic probing.

1.1 Norms and Submodularity

We start with the definitions of monotone, symmetric, and ordered norms. We will be only interested in norms defined in the positive orthant.

▶ **Definition 1** (Monotone Norm). *A* monotone norm *is a function* $\|\cdot\| : \mathbb{R}_+^n \to \mathbb{R}_+$ *and is defined as*

$$\|x\| := \sup_{\boldsymbol{a}\in\mathcal{A}} \sum_i a_i x_i,$$

i.e, by a max of non-negative linear functions over set $\mathcal{A}$*. This is equivalent to saying that* $\|x\| \geq \|y\|$ *whenever* $x \geq y \geq 0$ *coordinatewise (hence, the name* monotone*).*

▶ **Definition 2** (Symmetric Norm). *A monotone norm* $\|\cdot\|$ *is a* symmetric norm *if, for any vector* $x \in \mathbb{R}_+^n$ *and for all of its coordinate permutations* $\pi : [n] \to [n]$*,* $\|x\| = \|(x_{\pi(i)})_{i\in[n]}\|$*.*

We remark that any symmetric norm can be written as $\sup_{\boldsymbol{a}\in\mathcal{A}}\langle \boldsymbol{a}, x^{\downarrow}\rangle$, where $x^{\downarrow}$ represents the sorted (in descending order) vector x, and $\mathcal{A}$ is a set on non-negative descending vectors. This follows from the fact that $\max_\pi \sum_i a_i x_{\pi(i)} = \langle \boldsymbol{a}^{\downarrow}, x^{\downarrow}\rangle$ for non-negative vectors a, x. In the special case where $\mathcal{A}$ is a singleton, we have ordered norms.

▶ **Definition 3** (Ordered Norm). *A monotone norm is an* ordered norm *if it can be written as* $\|x\| = \sum_i a_i x_i^{\downarrow}$*, where* $a_1 \geq \ldots \geq a_n \geq 0$*.*

Submodular Norms

Submodular set functions and their applications to optimization have been extensively studied; see books [38, 22]. Intuitively, they capture the notion of decreasing marginal gains. Although submodular functions were originally defined for discrete settings, the notion has been generalized to arbitrary lattices, in particular to real vectors [8]. This leads to the following notion of continuous submodularity, which has found several applications in machine learning [12, 10] and will be crucial in our definition of submodular norms. We discuss standard properties of continuous submodularity in Appendix A.1.

▶ **Definition 4** (Continuous Submodularity). *A real-valued function $f : \mathbb{R}_+^n \to \mathbb{R}_+$ is* continuously submodular *if for all $x, y \in \mathbb{R}_+^n$, we have $f(x \vee y) + f(x \wedge y) \leq f(x) + f(y)$, where $x \vee y$ and $x \wedge y$ are the coordinate-wise max and min of x and y respectively.*

Our first contribution is to define the following natural class of submodular norms.

▶ **Definition 5** (Submodular Norm). *A monotone norm $\|\cdot\|$ is* submodular *if it is continuously submodular.*

Examples of submodular norms include all ℓ_p norms and Ordered norms (see Observation 12). Moreover, the following theorem proved in Section 2.2 shows that any symmetric norm can be approximated by a submodular norm.

▶ **Theorem 6.** *Any symmetric norm can be $O(\log \rho)$ approximated by a submodular norm, where $\rho := \frac{\|(1,1,\ldots,1)\|}{\|(1,0,0,\ldots,0)\|} \leq n$. This approximation factor is tight up to $O(\log\log \rho)$ terms.*

There is an intimate connection between submodular norms and submodular set functions. Given a submodular norm $\|\cdot\|$, the set function $f : 2^{[n]} \to \mathbb{R}^+$ by $f(S) := \|\mathbf{1}_S\|$ is submodular, so every submodular norm is an extension of a submodular function. Moreover, if f is a monotone submodular function with $f(\emptyset) = 0$, then f can be extended to a submodular norm $\|\cdot\|$ by the *Lovász extension*:

$$\|x\| := \int_0^\infty f(\{j : t \leq x_j\})dt.$$

This observation that every submodular set function induces a continuously submodular norm via its Lovász extension has appeared several times before [9, 8]. However, our definition of submodular norms can capture many more natural norms. E.g., all ℓ_p norms are submodular but for $1 < p < \infty$ they cannot be written as a Lovász extension of a submodular set function since the dual-norm unit ball has an infinite number of vertices.

▶ Remark 7. A commonly studied variant of continuous submodularity is DR-submodularity [13, 20, 37]: a function $f : \mathbb{R}_+^d \to \mathbb{R}_+$ is *DR-submodular* if it satisfies *diminishing returns* meaning $f(w + ae_i) - f(w) \leq f(x + ae_i) - f(x)$ for all $x, w \in \mathbb{R}_+^d$ with $x \leq w$, $i \in [d]$, and $a \geq 0$. It is known that continuous submodularity is equivalent to having this diminishing returns inequality only when $x_i = w_i$; hence continuous submodularity is a weaker property. The class of DR submodular functions turns out to be uninteresting when looking at norms since the only DR-submodular norm is the ℓ_1-norm (up to rescaling coordinate-wise). See Appendix A.1 for proofs.

1.2 Applications

In addition to being a natural class of norms, submodular norms find multiple applications. We will explore two specific applications-one in the domain of online algorithms and another in the field of stochastic optimization.

Online Facility Location

In this problem we are given a metric space $(\mathcal{M}, d)$ equipped with metric $d : \mathcal{M} \times \mathcal{M} \to \mathbb{R}_{\geq 0}$, along with a cost function $f : \mathcal{M} \to \mathbb{R}_+$ and a norm $\|\cdot\| : \mathbb{R}_+^n \to \mathbb{R}_+$. At each time step $i \in [n]$, an adversary produces a new request $x_i \in \mathcal{M}$ and the algorithm decides to either assign x_i to the closest already-open facility in the set F_{i-1}, thereby incurring a connection cost $d(x_i, F_{i-1})$, or to open a new facility q and assign request x_i to facility q, thereby

incurring a connection cost $d(x_i, q)$ and an opening cost $f(q)$. Let F be the final set of opened facilities, let F_i be the set of facilities opened until (and including) the i-th request, and let $\boldsymbol{d} = (d_1, \ldots, d_n) \in \mathbb{R}_+^n$ be the vector of connection costs $d_i := d(x_i, F_i)$. Our goal is to minimize the total cost $\sum_{q \in F} f(q) + \|\boldsymbol{d}\|$.

Online facility location was introduced by Meyerson for ℓ_1 norm [35], where he showed an $O(\log n)$ competitive algorithm. A tight competitive ratio of $\Theta(\log n / \log\log n)$ was later obtained by Fotakis [21]. When all requests are given up front (offline setting), it is a classical NP-hard problem where we can design $O(1)$ approximation algorithm, even for general norms [27]. In the online setting, however, no non-trivial algorithm was previously known beyond ℓ_1 norms.

▶ **Theorem 8.** *For online facility location problem with a submodular norm $\|\cdot\|$, there exists a randomized online algorithm that obtains cost at most $O(\log \rho) \cdot \sum_{z \in F^*} f(z) + O(1) \cdot \|\boldsymbol{d}^*\|$, where F^* and $\boldsymbol{d}^*$ are the set of facilities and vector of assignment distances respectively given by the offline optimum algorithm and $\rho := \frac{\|(1,1,\ldots,1)\|}{\min_i \|e_i\|} \leq n \cdot \frac{\max_i \|e_i\|}{\min_i \|e_i\|}$.*

Since any symmetric norm can be $O(\log \rho)$ approximated by a submodular norm by Theorem 6, we get the following corollary.

▶ **Corollary 9.** *For online facility location problem with a symmetric norm, there exists an $O(\log \rho)$-competitive randomized algorithm.*

For concreteness, this corollary implies an $O(\log n)$-competitive algorithm for ℓ_1 norm, which matches Meyerson's bound [35], an $O(1)$-competitive algorithm for ℓ_∞ norm, and an $O(\log k)$-competitive algorithm for Top-k norm. This is tight up to an $O(\log\log \rho)$ factor for any symmetric norm by the lower bound construction given in Theorem 31.

The proof of Theorem 8 relies on generalizing Meyerson's algorithm beyond ℓ_1 norms. Meyerson's algorithm constructs a new facility at each demand point x_i with probability $d(x_i, F_{i-1})/f$, thereby balancing the cost of assigning the demand against the cost of constructing a new facility. A natural generalization of this algorithm to general norms is to construct a new facility with probability δ_i/f, where marginal cost $\delta_i = \|(d_1, \ldots, d_{i-1}, d(x_i, F_{i-1}), 0, \ldots, 0)\| - \|\boldsymbol{d}_{\leq i-1}\|$. Unfortunately, we will show that such an algorithm is $\Omega(n)$-competitive even for the ℓ_∞ norm. Our crucial change to Meyerson's algorithm is to carefully define *auxiliary assignment costs* $\hat{d}_i$ which upper bound the true costs d_i. Now we use $\hat{d}_i$ instead of d_i to calculate the marginal cost δ_i. Due to norm submodularity, this underestimates the marginal costs, making the algorithm more inclined to assign demand points instead of constructing new facilities.

Next, we discuss a stochastic optimization application of submodular norms.

Stochastic Probing

This problem is a natural stochastic generalization of constrained submodular maximization. Here, we are given probability distributions of n independent random variables $X = (X_1, \ldots, X_n) \in \mathbb{R}_+$, a downward-closed set family $\mathcal{F} \subseteq 2^{[n]}$, and a monotone objective $f : \mathbb{R}_+^n \to \mathbb{R}_+$. The goal is to select a feasible set $S \in \mathcal{F}$ of variables in order to maximize $f(X_S)$. The optimal strategy for this problem is generally *adaptive*, i.e., it selects elements of S one at a time and may change its decisions based on observations of the selected variables.

Since adaptive strategies are complicated (could be an exponential-sized decision tree) and hard to implement for many applications of stochastic probing, we are interested in finding non-adaptive algorithms that maximize $\max_{S \in \mathcal{F}} \mathbb{E}[f(X_S)]$. The main question, which has been studied in several papers [5, 24, 25, 26, 14, 19], is how much do we lose when we

move from adaptive to non-adaptive algorithms, i.e., if $\mathsf{Adap}(X, \mathcal{F}, f)$ denotes the optimal adaptive strategy and $\mathsf{NA}(X, \mathcal{F}, f)$ denotes the optimal non-adaptive algorithm, then what is the maximum possible *adaptivity gap* $\frac{\mathsf{Adap}(X,\mathcal{F},f)}{\mathsf{NA}(X,\mathcal{F},f)}$.

For submodular set functions, the worst-case adaptivity gap is known to be 2 [26, 14]. An interesting conjecture posed in [26] is whether the adaptivity gap for *XOS* set functions is poly-logarithmic in n, where an XOS set function $f : 2^{[n]} \to \mathbb{R}_+$ is a max over linear set functions. Since a monotone norm is nothing but a max over linear functions (given by the dual-norm unit ball), they form an extension of XOS set functions from the hypercube to all non-negative real vectors. Thus, we can generalize the conjecture of [26] to the following:

▶ **Conjecture 10.** *The adaptivity gap for stochastic probing with monotone norms is poly-*$\log(n)$.

Although we are not able to resolve this general conjecture, we make progress by resolving it for all symmetric norms.

▶ **Theorem 11.** *The adaptivity gap for stochastic probing with symmetric norms is* $O(\log n)$.

The proof of this result relies on first approximating the symmetric norm by a submodular norm as given in Theorem 6. Next, we generalize the technique of bounding adaptivity gaps for submodular set functions in [14] to submodular norms.

1.3 Further Related Work

In recent years, there has been a surge of interest in the study of general norms. Some of the combinatorial problems that have been studied beyond ℓ_p norms are load balancing [16, 17, 29, 30], k-clustering [15, 16], vector scheduling [32, 18, 31], set cover [7, 36], spanning trees [28], and generalized assignment with convex costs [23, 32]. Beyond combinatorial optimization, general norms have been recently studied for problems such as mean estimation with statistical queries [34], nearest-neighbor search [4, 3], regression [2, 39], and communication complexity [1].

Continuous submodular functions have been extensively studied in the machine learning literature. We refer to the beautiful article of Bach [8] for their properties. Some of their applications to combinatorial optimization are discussed in [6, 37] and to learning are discussed in [40, 20]. The fact that submodular set function induces a norm via its Lovász extension has found several applications for regression since they induce sparsity [9, 10].

Paper Outline

Our work revolves around submodular norms and combinatorial optimization problems where the objective function is a submodular norm. In Section 2, we illustrate the key properties of submodular norms and the extent to which they serve as a good proxy for other classes of norms. In this respect, we identify a crucial parameter ρ that structurally characterizes a given submodular norm and may be of independent interest. In Section 3, we leverage these properties to derive a competitive algorithm for online facility location. Finally, in Section 4, we provide an application of submodular norms to adaptivity gaps for stochastic probing.

2 Submodular Norms

We study properties of submodular norms and how they relate to other commonly studied norms.

2.1 Properties and Important Special Cases

We first discuss some common examples of submodular norms.

▶ **Observation 12.** *The following norms are submodular:*
1. *All ℓ_p norms are submodular.*
2. *All Top-k and ordered norms are submodular.*
3. *For a matroid $\mathcal{M} = ([n], \mathcal{I})$, the* matroid rank norm $\|x\| := \max_{S \in \mathcal{I}}(\sum_{i \in S} x_i)$ *is submodular.*

Proof. To see that ℓ_p norms are submodular, it suffices to show that for any monotone concave $g : \mathbb{R}_+ \to \mathbb{R}_+$, and submodular $f : \mathbb{R}_+^n \to \mathbb{R}_+$, the function $g \circ f$ is submodular. We can then apply this when $f = \|x\|_p^p$ and $g(y) = y^{1/p}$.

To prove the claim, notice that for $x, y \in \mathbb{R}_+^n$,

$$g(f(x \vee y)) - g(f(x)) \leq g(f(x \vee y) - f(x) + f(x \wedge y)) - g(f(x \wedge y)) \leq g(f(y)) - g(f(x \wedge y)).$$

On the other hand, Top-k norms and matroid rank norms are special cases of Lovász extensions. The matroid rank norm is the Lovász extension of the rank function, and a Top-k norm is a matroid rank norm for the k-uniform matroid. ◀

Submodular norms are also closed under several natural operations.

▶ **Lemma 13.** *The following operations return a submodular norm:*
1. *Any rescaling of the coordinates of a submodular norm.*
2. *Sums of partial*[1] *submodular norms.*
3. *Any conical combination of submodular norms is submodular.*[2]

Proof. The first property follows since coordinate-wise rescaling of vectors commutes with coordinate-wise max and min.

The second property follows from the fact that a partial submodular norm is a submodular semi-norm (i.e., a norm without the requirement to be positive definite). A sum of semi-norms remains a semi-norm, and from [8], a sum of continuously submodular functions is continuously submodular.

Finally, it is folklore that conical combinations of norms are norms, and it is also easy to show that such combinations also preserve continuous submodularity (see [8]). ◀

Besides their strict containment of many common norms, submodular norms are also powerful because they can be used to approximate other norms. In Section 2.2, we will discuss how symmetric norms can be approximated by submodular norms up to logarithmic factors. In addition, we note in Section 2.3 that submodular norms may approximate a much larger class of norms than just symmetric, although they are still far from the most general class of monotone norms. These approximation relations are summarized in Figure 1.

---

[1] Partial means norms defined on a subset of coordinates with every other coordinate treated as 0.

[2] Let $x_1, \ldots, x_m \in \mathbb{R}^n$ be real-valued vectors. We say that $y = \sum_{i \in [m]} \alpha_i x_i$ is a *conical* combination of the vectors if $\alpha_i \geq 0$ for all $i \in [m]$.

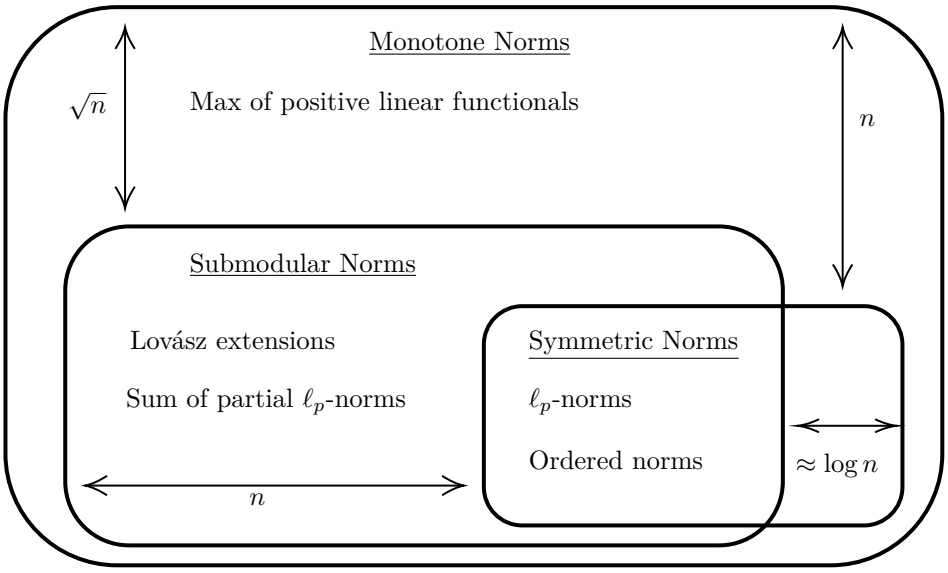

Figure 1 The containment relationships between monotone norms, submodular norms, and symmetric norms, along with some examples. The "distances" shown indicate the worst-case approximation factor (up to constants) for a norm in each outer class by a norm in the corresponding inner class.

2.2 Approximation of Symmetric Norms

A major benefit of studying submodular norms is that they can approximate any symmetric norm. Indeed, previous works have noted that symmetric norms can be approximated by an ordered norm up to a factor of $O(\log n)$ [16, 32]. For our purposes, it will be useful to make this approximation more precise by replacing $\log n$ with $\log \rho$, where parameter ρ is defined as follows:

▶ **Definition 14.** *If $e_1, \ldots, e_n$ denote the standard basis vector and let $\mathbf{1}_{\leq i} := \sum_{1 \leq j \leq i} e_j$ denote the vector with 1s at the first i coordinates and 0 otherwise. Then for any monotone norm $\|\cdot\|$ we define the parameter*

$$\rho_{\|\cdot\|} := \frac{\|\mathbf{1}_{\leq n}\|}{\min_{i \in [n]} \|e_i\|}.$$

When the norm is clear from context, we simply write $\rho = \rho_{\|\cdot\|}$.

Notice that for symmetric norms, we have $\rho = \frac{\|\mathbf{1}_{\leq n}\|}{\|e_1\|} \leq n$. One can think of ρ for symmetric norms as a measure of how closely a norm behaves like $\|\cdot\|_1$ (large ρ) versus $\|\cdot\|_\infty$ (small ρ).

▶ **Observation 15.** *For ℓ_p norms, we have $\rho_{\|\cdot\|_p} = n^{1/p}$. For Top-$k$ norms, we have $\rho_{\|\cdot\|_{\text{Top-}k}} = k$.*

As we will see, the parameter ρ appears again in both the upper and lower bound analysis in Section 3 and Appendix B.3, making the improvement from $\log n$ to $\log \rho$ in Lemma 16 necessary for tight bounds in our applications.

▶ **Lemma 16.** *For any symmetric norm $\|\cdot\|$ with $\rho_{\|\cdot\|} = \rho$, there is an ordered norm $\|\cdot\|'$ such that $\|x\| \leq \|x\|' \leq 2(\log \rho + 1) \cdot \|x\|$.*

Proof. Let $\|x\| = \max_{a\in\mathcal{A}}\langle a, x^{\downarrow}\rangle$. Without loss of generality, assume $\|e_1\| = 1$, so $\|\mathbf{1}_{\le n}\| = \|(1,\dots,1)\| = \rho$.

Let $1 = m_0 \le m_1 \le \cdots \le m_{\lfloor \log\rho\rfloor}$ be such that m_j is the least integer with $\|\mathbf{1}_{\le m_j}\| \ge 2^j$. Let $a_0,\dots,a_{\lfloor\log\rho\rfloor} \in \mathcal{A}$ be such that $\|\mathbf{1}_{\le m_j}\| = \langle a_j, \mathbf{1}_{\le m_j}\rangle$.

Now consider the ordered norm $\|x\|' := 2\langle a^*, x^{\downarrow}\rangle$, where $a^* = \sum_j a_j$. Clearly, we have

$$\frac{1}{2(\lfloor\log\rho\rfloor+1)}\|x\|' \le \max_j\langle a_j, x^{\downarrow}\rangle \le \max_{a\in\mathcal{A}}\langle a, x^{\downarrow}\rangle = \|x\|.$$

Additionally, notice that for any $x \in \mathbb{R}^n_+$, we can write $x^{\downarrow} = \sum_{k\in[n]}\lambda_k\mathbf{1}_{\le k}$ for some $\lambda_k \ge 0$. We have

$$\|x\|' = 2\sum_j\sum_k \lambda_k\langle a_j, \mathbf{1}_{\le k}\rangle \ge 2\sum_k\lambda_k\max_j\langle a_j,\mathbf{1}_{\le k}\rangle \ge^{\dagger} \sum_k\lambda_k\|\mathbf{1}_{\le k}\| \ge \|x\|.$$

Here, † follows from rounding each k down to the nearest m_j and using

$$\langle a_j, \mathbf{1}_{\le k}\rangle \ge \langle a_j, \mathbf{1}_{\le m_j}\rangle = \|\mathbf{1}_{\le m_j}\| \ge \frac{1}{2}\|\mathbf{1}_{\le k}\|. \qquad \blacktriangleleft$$

Tightness of approximation

In the worst case where $\rho = \Omega(n)$, the $\log n$ factor turns out to be nearly the best possible factor for approximation of a symmetric norm by a submodular norm. The following lemma shows that the construction in Lemma 16 is tight up to $O(\log\log n)$ factors.

▶ **Lemma 17.** *For any $\varepsilon \in (0, 1/2)$, define*

$$\|x\| := \max_{k\in[n]} k^{-\varepsilon}\cdot\langle \mathbf{1}_{\le k}, x^{\downarrow}\rangle.$$

For any submodular norm $\|\cdot\|'$ such that $\|x\|' \ge \|x\|$ for all $x \in \mathbb{R}^n_+$, there exists $y \in \mathbb{R}^n_+$ such that $\|y\|' \ge C\frac{\varepsilon}{1-\varepsilon}(\log n)^{1-\varepsilon}\|y\|$. Taking $\varepsilon = \frac{1}{\log\log n}$ gives $\|y\|' \ge \Omega(\frac{\log n}{\log\log n})\|y\|$.

Proof. Let y be defined by $y_k := \frac{k^{\varepsilon}-(k-1)^{\varepsilon}}{\varepsilon}$. A simple calculation yields that

$$k^{-(1-\varepsilon)} \le y_k \le (k-1)^{-(1-\varepsilon)}.$$

Additionally, we have

$$\|y\| = \max_{k\in[n]} k^{-\varepsilon}\cdot\sum_{i=1}^{k} y_i = \max_{k\in[n]} k^{-\varepsilon}\cdot\frac{k^{\varepsilon}}{\varepsilon} = \frac{1}{\varepsilon}.$$

Now to estimate $\|y\|'$, first note that we may assume $\|\cdot\|'$ to be symmetric, otherwise replace $\|\cdot\|'$ with its average over all permutations of inputs. We will inductively show that for $j \in [n/\log n]$, we have $\|y_{\le j}\|' \ge b_j := \frac{(\log(j+1))^{1-\varepsilon}}{4(1-\varepsilon)}$.

For $j = 1$, we check

$$b_1 \le \frac{1}{4(1-\varepsilon)} \le \frac{1}{2} \le 1 = \|y_{\le 1}\| \le \|y_{\le 1}\|'.$$

Now assume the claim holds for a given $j \ge 1$. Consider $z \in \mathbb{R}^d_+$ defined by

$$z_i = \begin{cases} y_i & 1 \le i \le j, \\ y_{j+1} & j < i \le j+\ell, \\ 0 & j+\ell < i, \end{cases}$$

where $\ell := \lceil (j+1)\log(j+1) \rceil$. Notice that by submodularity and symmetry, we have

$$\|y_{\leq j+1}\|' \geq \|y_{\leq j}\|' + \frac{\|z\|' - \|y_{\leq j}\|'}{\ell} \geq b_j + \frac{\|z\|' - b_j}{\ell}.$$

We see that

$$\|z\|' \geq \|z\| \geq (j+\ell)^{-\varepsilon} \cdot (j+\ell) y_{j+1} \geq \left(\frac{j+\ell}{j+1}\right)^{1-\varepsilon} \geq \log(j+1)^{1-\varepsilon}.$$

Thus, we have $\|z\|' - b_j \geq \frac{1}{2}\log(j+1)^{1-\varepsilon}$. Finally, we have

$$\|y_{\leq j+1}\|' \geq b_j + \frac{\log(j+1)^{1-\varepsilon}}{2\ell} \geq b_j + \frac{\log(j+1)^{-\varepsilon}}{4(j+1)} \geq b_j + \int_{j+1}^{j+2} \frac{(\log x)^{-\varepsilon}}{4x} dx = b_{j+1}. \quad \blacktriangleleft$$

2.3 Beyond Symmetric Norms

Given that submodular norms allow us to approximate symmetric norms up to $\log n$ factors, we may ask if other classes of norms can be similarly approximated. We note that there exist submodular norms that are an $\Omega(n)$ factor away from any symmetric norm, which suggests that symmetric norms are not the largest class of norms which are approximated by submodular norms. Indeed, sums of partial ℓ_p or Top-k norms, such as those considered in [36], are submodular but can be highly asymmetric. Thus, although we focus on approximating symmetric norms by submodular norms, it is likely that many asymmetric norms admit submodular approximations. However, we leave the problem of characterizing these norms for future work.

In the following lemmas, we adopt the notation x_S for $x \in \mathbb{R}^n$ and $S \subseteq [n]$ to denote x after zeroing out all entries except those at indices in S, as well as $\mathbf{1}_S$ to denote the indicator vector of S.

▶ **Lemma 18.** *There exists a submodular norm $\|\cdot\|'$ for which any symmetric norm $\|\cdot\|$ satisfying $\|x\| \leq \|x\|'$ for all $x \in \mathbb{R}^n_+$, also has $\|y\|' \geq \Omega(n) \cdot \|y\|$ for some $y \in \mathbb{R}^n_+$.*

Proof. Let $A := \{1, \ldots, n/2\}$ and $B := \{n/2 + 1, \ldots, n\}$. Define the norm $\|x\|' := \|x_A\|_\infty + \|x_B\|_1$. Notice that $\|\cdot\|'$ is a sum of partial ℓ_p norms, so it is submodular by Lemma 13. Now suppose $\|\cdot\|$ is a symmetric norm with $\|x\| \leq \|x\|'$ for all $x \in \mathbb{R}^n_+$. Then $\|\mathbf{1}_A\| \leq \|\mathbf{1}_A\|' = 1$. However, we also have $\|\mathbf{1}_B\| = \|\mathbf{1}_A\|$ by symmetry, and $\|\mathbf{1}_B\|' = n/2$. Thus, taking $y = \mathbf{1}_B$, we have our lemma. ◀

In the most general setting of monotone norms, however, submodular norms cannot give better than an $\Omega(\sqrt{n})$ approximation. The proof of this fact is similar to the canonical proof of the $\Omega(\sqrt{n})$ factor gap between submodular set functions and XOS set functions.

▶ **Lemma 19.** *There exists a monotone norm $\|\cdot\|$ for which any submodular norm $\|\cdot\|'$ satisfying $\|x\| \leq \|x\|'$ for all $x \in \mathbb{R}^n_+$, also has $\|y\|' \geq \Omega(\sqrt{n}) \cdot \|y\|$ for some $y \in \mathbb{R}^n_+$.*

Proof. Partition $[n]$ into $\sqrt{n}$ blocks $B_1, \ldots, B_{\sqrt{n}}$, each of size $\sqrt{n}$. Define the norm $\|x\| := \max_{k \in [\sqrt{n}]} \left(\sum_{i \in B_k} x_i\right)$. Now suppose $\|\cdot\|'$ is a submodular norm satisfying $\|x\| \leq \|x\|'$ for all $x \in \mathbb{R}^n_+$. We will construct y by starting with the zero vector and iteratively choosing one element i_k of each B_k to activate (set $y_{i_k} = 1$). Clearly $\|y\| = 1$, so we just need to show $\|y\|' \geq \Omega(\sqrt{n})$.

Formally, let $y^{(0)} = 0$, and for each $k = 1, \ldots, \frac{\sqrt{n}}{2}$, do the following. If $\|y^{(k-1)}\|' \geq \frac{\sqrt{n}}{2}$, we are done and simply choose $y = y^{(k-1)}$. Otherwise, notice that $\|y^{(k-1)} + \mathbf{1}_{B_k}\|' \geq \|y^{(k-1)} + \mathbf{1}_{B_k}\| = \sqrt{n}$, so $\|y^{(k-1)} + \mathbf{1}_{B_k}\|' - \|y^{(k-1)}\| \geq \frac{\sqrt{n}}{2}$. By submodularity, there exists some $i_k \in B_k$ such that $\|y^{(k-1)} + e_{i_k}\|' \geq \|y^{(k-1)}\|' + \frac{1}{2\sqrt{n}}$, for which we set $y^{(k)} := y^{(k-1)} + e_{i_k}$ By induction, if we do not terminate early, we have $\|y^{(\sqrt{n})}\|' \geq \frac{n}{2\sqrt{n}} = \frac{\sqrt{n}}{2}$. ◀

3 Online Facility Location with Submodular Norms

In this section, we illustrate how submodular norms can be applied to Online Facility Location. Recall from Section 1.2, in this problem we are given a metric space $(\mathcal{M}, d)$ along with a cost function $f : \mathcal{M} \to \mathbb{R}_+$. At each time step $i \in [n]$, an adversary produces a new request $x_i \in \mathcal{M}$, and the algorithm decides whether to assign x_i to the closest open facility F_{i-1} or to open a new facility q and assign request x_i to q. The goal is to minimize the total facility opening costs plus a given norm $\|\cdot\|$ of the connection costs, i.e., $\min \sum_{q \in F} f(q) + \|\boldsymbol{d}\|$, where $\boldsymbol{d} = (d_1, \ldots, d_n) \in \mathbb{R}_+^n$ is the vector of connection costs $d_i := d(x_i, F_i)$.

In the case of uniform costs, $f(q) = f$ for all q, so the total facility opening cost becomes $f \cdot |F|$. We use $\boldsymbol{d}_{\leq i} = (d_1, \ldots, d_i, 0, \ldots, 0)$ to denote the first i coordinates of vector $\boldsymbol{d}$.

3.1 Uniform Costs

For now, we will focus on the case when facility costs are uniformly f.

▶ **Theorem 20.** *Let $\|\cdot\|$ be a submodular norm, and let $\rho := \rho_{\|\cdot\|}$. For the $\|\cdot\|$ norm online facility location problem with uniform facility costs f, there exists a randomized online algorithm that obtains cost at most $O(\log \rho) \cdot |F^*| f + O(1) \cdot \|\boldsymbol{d}^*\|$, where F^* and $\boldsymbol{d}^*$ are the set of facilities and vector of assignment distances, respectively, given by the optimal offline algorithm.*

Notice that because our algorithm obtains a constant factor approximation for the assignment costs, we have the following corollary.

▶ **Corollary 21.** *There is an $O(\log \rho)$-competitive algorithm for uniform costs online facility location with symmetric norms.*

Proof. Given the uniform cost facility location problem with a monotone symmetric norm $\|\cdot\|$, let F^* and $\boldsymbol{d}^*$ be the set of facilities and assignments distances given by the optimal offline algorithm. For our online algorithm, we will approximate $\|\cdot\|$ by a submodular norm $\|\cdot\|'$ using Lemma 16, and then run the algorithm in Theorem 20 on norm $\|\cdot\|'$. Since $\log \rho = \log \rho_{\|\cdot\|} = \Theta(\log \rho_{\|\cdot\|'})$, this algorithm will incur cost at most

$$|F|f + \|\boldsymbol{d}\| \leq |F|f + \|\boldsymbol{d}\|' \leq O(\log \rho)|F^*|f + O(1)\|\boldsymbol{d}^*\|' \leq O(\log \rho)|F^*|f + O(\log \rho)\|\boldsymbol{d}^*\|. \blacktriangleleft$$

Proof outline for Theorem 20

We want to generalize Meyerson's algorithm beyond ℓ_1 norms and use submodularity to complete the analysis. Meyerson's algorithm constructs a new facility at each demand point x_i with probability $d(x_i, F_{i-1})/f$, thereby balancing the cost of assigning the demand against the cost of constructing a new facility. To adapt this algorithm to more general norms, it is natural to construct a new facility at x_i with probability δ_i/f, where $\delta_i = \|(d_1, \ldots, d_{i-1}, d(x_i, F_{i-1}), 0, \ldots, 0)\| - \|\boldsymbol{d}_{\leq i-1}\|$ is the marginal cost of assigning x_i.

Unfortunately, the above natural generalization of Meyerson's algorithm can have an $\Omega(n)$ competitive ratio. For instance, consider the star graph $K_{1,n}$ equipped with the standard unweighted graph distance metric. Suppose that our construction costs are $f = 1$, our submodular norm is the ℓ_∞ norm, and the demand points are all the leaves of the star graph. The optimal solution constructs a single facility at the center, yielding a total cost of $1 + \|(1, \ldots, 1)\|_\infty = 2$. On the other hand, the suggested algorithm constructs a facility for every demand point, as $\delta_i = 2$ for each $i \geq 2$, incurring a total cost of n.

To get around this issue, we will define *auxiliary assignment costs* $\hat{d}_i$ that upper bound the true costs d_i. The key modification to the algorithm is that we will use $\hat{d}_i$ instead of d_i for calculating the marginals δ_i. By overestimating the assignment costs that we have incurred, the algorithm underestimates potential marginal costs due to submodularity, making it more inclined to assign demand points instead of constructing new facilities. Moreover, the flexibility of the analysis allows us to show that the increased costs of $\hat{d}_i$ still obtain an $O(\log \rho)$ competitive ratio.

We now present the formal proof.

Proof of Theorem 20. To formalize the outline above, we will first inductively define our auxiliary cost vector $\hat{\boldsymbol{d}}$ and the marginals δ_i by

$$\hat{d}_i := \min\left\{d(x_i, F_{i-1}),\ \ \max\{z \geq 0 : f \geq \|(\hat{d}_1, \ldots, \hat{d}_{i-1}, z, 0, \ldots, 0)\| - \|\hat{\boldsymbol{d}}_{\leq i-1}\|\}\right\} \quad \text{and}$$
$$\delta_i := \|\hat{\boldsymbol{d}}_{\leq i}\| - \|\hat{\boldsymbol{d}}_{\leq i-1}\|.$$

Thus, $\hat{d}_i$ is the assignment distance $d(x_i, F_{i-1})$ capped such that $\delta_i \leq f$. For our algorithm, we construct a facility at x_i with probability δ_i/f, and assign x_i to the nearest facility otherwise. These are well-defined probabilities since $\delta_i \leq f$. To see the upper-bound $d_i \leq \hat{d}_i$, notice that if $\delta_i < f$, then $\hat{d}_i = d(x_i, F_{i-1}) \geq d(x_i, F_i) = d_i$. If $\delta_i = f$, then $d_i = 0$ since a facility is constructed at x_i with probability 1, so $d_i = 0 \leq \hat{d}_i$.

Let $\mathsf{cost}(i) := f \cdot \mathbb{1}_{|F_i| > |F_{i-1}|} + \delta_i$ be the marginal increase in auxiliary cost at step i, so $\sum_{i \in [n]} \mathsf{cost}(i) = f \cdot |F| + \|\hat{\boldsymbol{d}}\|$. We will bound separately the cost of demand points that arrive before and after the first nearby facility is constructed. Similar to Meyerson's proof, we have the following bound on costs incurred before a facility is constructed in a given set.

▷ Claim 22. Let $A \subseteq [n]$ be a fixed set of indices, and let $S \subseteq A$ be the subset of indices that arrive before the first facility is constructed at any step in A. Then $\mathbb{E}[\sum_{i \in S} \mathsf{cost}(i)] \leq 2f$.

The proof of this claim is essentially identical to Meyerson's, so we will defer it to Appendix B.1.

Now, let us enumerate our offline algorithm's facility set as $F^* = \{c_1^*, \ldots, c_K^*\}$, where $K = |F^*|$. Let $C_1^*, \ldots, C_K^*$ be the offline clusters, i.e., C_k^* is the set of $i \in [n]$ for which x_i is assigned to c_k^*.

Let $r := \frac{\|\boldsymbol{d}^*\|}{\|(1,\ldots,1)\|}$. We partition each cluster into rings as $C_k^* = \bigcup_{\ell=0}^{L} C_k^\ell$, where $L = \lceil \log \rho \rceil$,

$$C_k^0 := \{i \in C_k^* : d(x_i, c_k^*) \leq r\}, \quad \text{and}$$
$$C_k^\ell := \{i \in C_k^* : 2^{\ell-1} r \leq d(x_i, c_k^*) \leq 2^\ell r\} \quad \text{for } \ell \in \{1, \ldots, L\}.$$

Notice that this is a partition since

$$\max_{i \in C_k^*} d(x_i, C_k^*) \leq \|\boldsymbol{d}^*\|_\infty \leq \frac{\|\boldsymbol{d}^*\|}{\min_i \|e_i\|} = r\rho \leq 2^L r.$$

We will analyze the costs incurred by our algorithm on demand points in each ring. Within each ring, we will consider two types of demands separately: *long-distance* demands LD_k^ℓ and *short-distance* demands SD_k^ℓ, defined as

$$\mathsf{LD}_k^\ell := \{i \in C_k^\ell : d(c_k^*, F_{i-1}) > 2^\ell r\} \quad \text{and} \quad \mathsf{SD}_k^\ell := \{i \in C_k^\ell : d(c_k^*, F_{i-1}) \le 2^\ell r\}.$$

In other words, long (respectively, short) distance demands arrive before (respectively, after) a facility has been constructed within the outer perimeter of its corresponding ring. We now make the following claims.

▷ Claim 23. We have

$$\mathbb{E}\Big[\sum_{\ell=0}^{L}\sum_{k=1}^{K}\sum_{i\in\mathsf{LD}_k^\ell}\mathsf{cost}(i)\Big] \le 2(L+1)Kf.$$

▷ Claim 24. We have

$$\mathbb{E}\Big[\sum_{\ell=0}^{L}\sum_{k=1}^{K}\sum_{i\in\mathsf{SD}_k^\ell}\mathsf{cost}(i)\Big] \le 8\|\boldsymbol{d}^*\|.$$

These claims together give Theorem 20, since $L = O(\log\rho)$.

Notice that Claim 23 follows immediately from Claim 22 since long-distance demands must arrive before a facility is constructed in $C_\ell^{(j)}$. This implies $\mathbb{E}\left[\sum_{i\in\mathsf{LD}_k^\ell}\mathsf{cost}(i)\right] \le 2f$ for each ℓ and k.

To show Claim 24, let $\mathsf{SD} := \bigcup_{\ell=0}^{L}\bigcup_{k=1}^{K}\mathsf{SD}_k^\ell$. We seek to show that $\sum_{i\in\mathsf{SD}}\mathsf{cost}(i) \le 8\|\boldsymbol{d}^*\|$. Notice that if $i \in \mathsf{SD}_k^0$ for some k, then we have $\hat{d}_i \le d(x_i, F_{i-1}) \le d(x_i, c_k^*) + d(c_k^*, F_{i-1}) \le d_i^* + r$. Similarly, if $i \in \mathsf{SD}_k^\ell$ for some k and $\ell \ne 0$, we have $\hat{d}_i \le 3d_i^*$. Thus, we can say that $\hat{d}_i \le 3d_i^* + r$ for all $i \in \mathsf{SD}$. This gives

$$\mathbb{E}\Big[\sum_{i\in\mathsf{SD}}\mathsf{cost}(i)\Big] \;\le\; \mathbb{E}\Big[\sum_{i\in\mathsf{SD}}2\delta_i\Big] \;\le\; 2\mathbb{E}\Big[\|\hat{\boldsymbol{d}}_{\mathsf{SD}}\|\Big] \;\le\; 2\|r\mathbf{1}_{\le n}+3\boldsymbol{d}^*\| \;=\; 2r\|\mathbf{1}_{\le n}\|+6\|\boldsymbol{d}^*\| \;\le\; 8\|\boldsymbol{d}^*\|,$$

where the second inequality comes from the submodular property, the third comes from norm monotonicity, and the last inequality is by choice of r. In particular, for the second inequality, we crucially use the submodularity of the norm to say

$$\sum_{i\in\mathsf{SD}}\delta_i \;=\; \sum_{i\in\mathsf{SD}}\Big(\|\hat{\boldsymbol{d}}_{\le i}\| - \|\hat{\boldsymbol{d}}_{\le i-1}\|\Big) \;\le\; \sum_{i\in\mathsf{SD}}\Big(\|\hat{\boldsymbol{d}}_{\mathsf{SD}\cap[i]}\| - \|\hat{\boldsymbol{d}}_{\mathsf{SD}\cap[i-1]}\|\Big) \;=\; \|\hat{\boldsymbol{d}}_{\mathsf{SD}}\|. \qquad ◀$$

3.2 Non-Uniform Costs

In this section, we argue how one could develop the ideas of the previous section further to extend the algorithm to Online Facility Location with different opening costs across facilities. We will show how to modify Meyerson's algorithm for non-uniform costs in a similar manner to the uniform cost setting, but in a way that also handles new challenges that arise. We motivate and describe the new algorithm, but to avoid clutter, we defer most of the proof details to Appendix B.2.

▶ **Theorem 8.** *For online facility location problem with a submodular norm $\|\cdot\|$, there exists a randomized online algorithm that obtains cost at most $O(\log\rho)\cdot\sum_{z\in F^*}f(z) + O(1)\cdot\|\boldsymbol{d}^*\|$, where F^* and $\boldsymbol{d}^*$ are the set of facilities and vector of assignment distances respectively given by the offline optimum algorithm and $\rho := \frac{\|(1,1,\ldots,1)\|}{\min_i\|e_i\|} \le n\cdot\frac{\max_i\|e_i\|}{\min_i\|e_i\|}$.*

Recall that previously, the algorithm's only choice at step i was whether to construct a facility at x_i or not. In the non-uniform setting, it might not be feasible to only ever construct at x_i, since the cost of opening there might be prohibitively high. Meanwhile a nearby location might have a much lower cost. Instead, the algorithm must consider all possible cost levels at which it could construct and how far facilities at different cost levels are from x_i.

First, let us recap how Meyerson's algorithm handles this in the ℓ_1 norm setting. By losing at most a factor of 2, we can assume that all opening costs are in some set $\{f_1, \ldots, f_m\}$, where each f_j is a power of 2. For each $j \in [m]$, we define the $W_i^{(j)}$ to be the set of facilities which at step i are open or have opening cost at most f_j:

$$W_i^{(j)} := F_{i-1} \cup \{x \in \mathcal{M} : f(x) \leq f_j\}.$$

Additionally, $W_i^{(0)} := F_{i-1}$. Now Meyerson's algorithm [35] will, for each $j \in [m]$, open a facility at the nearest location in W_j with probability $\frac{d(x_i, W_i^{(j-1)}) - d(x_i, W_i^{(j)})}{f_j}$, capped at 1, then assign x_i to the nearest open facility. As in the uniform case, we can see that the expected facility opening cost incurred is $d(x_i, W_i^{(0)}) = d(x_i, F_{i-1})$. This allows us to again consider a "long-distance" and "short-distance" phase within each ring of each optimal cluster.

To adapt this algorithm to the submodular norm setting, the idea will again be to consider construction probabilities, not based on the distances $d(x_i, W_i^{(j)})$, but instead based on the marginal increase in an auxiliary assignment cost $\|\hat{\boldsymbol{d}}\|$. However, the definition of $\hat{\boldsymbol{d}}$ is not as straightforward as in the uniform cost setting. In particular, we need to satisfy $d_i \leq \hat{d}_i$ for each i deterministically. However, without carefully controlling dependencies between facility construction (for instance, by constructing independently for each j), it may be possible that $d_i = d(x_i, F_{i-1})$ if no facilities are constructed. This would force $\hat{d}_i$ to be too large to serve as a useful upper bound.

To avoid this, we instead construct facilities by sampling a single cost level f_j at which to build a facility. The probability of sampling f_j is given by a version of the probabilities used above. Suppose these probabilities sum to a value greater than 1. In that case, we crucially limit the probabilities for smaller f_j (corresponding to a facility at a greater distance) until the total probability is capped at 1: We name these probabilities $p_i^{(j)}$'s (Definition 30). This ensures that the true assignment distance d_i is never too large, which allows us to pick a conveniently small upper bound $\hat{d}_i$.

Algorithm 1 Non-Uniform Submodular Online Facility Location.

Data: Metric space $(\mathcal{M}, d)$ and online requests $x_1, \ldots, x_n$
Result: Opened facilities set F and assignment of requests to facilities (at the time of arrival)
for *request* x_i, $i = 1, \ldots, n$ **do**
  Sample $j \in \{0, \ldots, m\}$ according to the distribution given by $p_i^{(j)}$;
  Assign x_i to the nearest location in $W_i^{(j)}$, constructing a facility there if necessary;
end

Our analysis proceeds by considering the partition of arrival indices into optimal clusters $[n] = \bigcup_{k \in [K]} C_k$, each with center c_k^*. Similarly to the uniform cost setting, we partition each cluster into rings as $C_k^* = \bigcup_{\ell=0}^{L} C_k^\ell$, where $L = \lceil \log \rho \rceil$ and

$$C_k^0 := \{i \in C_k^* : d(x_i, c_k^*) \leq r\},$$
$$C_k^\ell := \{i \in C_k^* : 2^{\ell-1} r \leq d(x_i, c_k^*) \leq 2^\ell r\}, \quad \text{for } \ell \in [L].$$

For each ring C_k^ℓ, we divide the analysis into two main stages: the *short-distance* and *long-distance* stages. We leave the formal definition of these stages for the proof details, but intuitively, each demand is considered a long-distance demand until a facility is constructed at a distance which is a constant multiple of the radius of C_k^ℓ. After this happens, subsequent demands in the ring are considered short-distance demands. In each of these stages, the algorithm obtains the following.

▶ **Lemma 25** (Short-Distance Stage). *The expected cost incurred by the algorithm in the short-distance stage is*

$$\mathbb{E}\left[ALG_{SD}\right] = \mathbb{E}\Big[\sum_{i \in SD} cost(i)\Big] \leq 36\|\boldsymbol{d}^*\|.$$

▶ **Lemma 26** (Long-Distance Stage). *The expected cost incurred in the long-distance stage is*

$$\mathbb{E}\Big[ALG_{LD}\Big] \leq 48(\log \rho + 1) \cdot \sum_{k \in [K]} f(c_k^*).$$

It is now easy to see that Theorem 8 directly follows by combining Lemma 25 and Lemma 26. The proofs of these lemmas, along with a more formal definition of the probabilities $p_i^{(j)}$, are given in Appendix B.2.

4 Adaptivity Gaps for Stochastic Probing

In this section, we use submodular norms to prove small adaptivity gaps for the stochastic probing problem. Recall the stochastic probing problem from Section 1.2: Given n independent random variables $X = (X_1, \ldots, X_n) \in \mathbb{R}_+$, a downward closed set family $\mathcal{F} \subseteq 2^{[n]}$, and a monotone objective $f : \mathbb{R}_+^n \to \mathbb{R}_+$, the stochastic probing problem $(X, \mathcal{F}, f)$ is to open a feasible set $S \in \mathcal{F}$ of variables to maximize $f(X_S)$.

Denote by $\mathsf{Adap}(X, \mathcal{F}, f)$ the maximum expected objective achievable by an adaptive algorithm, i.e., one which selects elements of S one at a time, and may change its strategy based on its observations of the selected variables. We denote by $\mathsf{NA}(X, \mathcal{F}, f)$ the maximum expected objective by a non-adaptive algorithm, i.e., $\mathsf{NA}(X, \mathcal{F}, f) := \max_{S \in \mathcal{F}} \mathbb{E}[f(X_S)]$.

▶ **Theorem 27.** *If f is a submodular norm, then $Adap(X, \mathcal{F}, f) \leq 2 \cdot NA(X, \mathcal{F}, f)$.*

The following result for symmetric norms is an immediate corollary due to Theorem 6.

▶ **Theorem 11.** *The adaptivity gap for stochastic probing with symmetric norms is $O(\log n)$.*

Proof of Theorem 27. We follow the same proof approach as in [14]. Consider an adaptive algorithm Adap, and a non-adaptive algorithm Alg which selects each $S \in \mathcal{F}$ with the same probabilities as Adap, only non-adaptively. We will show by induction on n that Adap achieves an expected objective at most twice that of Alg. This is trivially true for $n = 1$, so we only need to show the inductive step.

We can compare the performance of these two algorithms by coupling their actions. Let's say Adap runs on random variables $X = (X_1, \ldots, X_n)$, and let $S \in \mathcal{F}$ be the (random) set of adaptively chosen variables. We can say Alg runs on variables $Y = (Y_1, \ldots, Y_n)$, i.i.d. copies of X, by choosing the same set S as Adap. Without loss of generality, say that Adap starts by selecting X_1. Since $1 \in S$ deterministically, we have that Adap achieves reward

$$\begin{aligned}
\mathbb{E}[f(X_S)] &= \mathbb{E}[f(X_1)] + \mathbb{E}\big[f_{X_1}(X_{S \setminus \{1\}}) \mid X_1\big] \\
&\leq \mathbb{E}[f(X_1 \vee Y_1)] + \mathbb{E}\big[f_{X_1 \vee Y_1}(X_{S \setminus \{1\}}) \mid X_1\big] \\
&\leq 2\mathbb{E}[f(Y_1)] + \mathbb{E}\big[f_{X_1 \vee Y_1}(X_{S \setminus \{1\}}) \mid X_1, Y_1\big],
\end{aligned}$$

where $f_x(Z) := f(x, Z) - f(x, 0)$ for $Z = (Z_2, \ldots, Z_n) \in \mathbb{R}_+^{n-1}$. Notice that, since f is submodular, we have f_x is submodular and decreasing in x for all $x \in \mathbb{R}_+$.

Now, notice that Alg achieves reward

$$\begin{aligned}\mathbb{E}[f(Y_S)] &= \mathbb{E}[f(Y_1)] + \mathbb{E}\big[f_{Y_1}(Y_{S\setminus\{1\}}) \mid X_1\big] \\ &\geq \mathbb{E}[f(Y_1)] + \mathbb{E}\big[f_{X_1 \vee Y_1}(Y_{S\setminus\{1\}}) \mid X_1, Y_1\big].\end{aligned}$$

Notice that given X_1, the set $S \setminus \{1\} \in \mathcal{F}|_{-1}$ is adaptively chosen among the variables $X_2, \ldots, X_n$ by Adap. Thus, by induction we can say $\mathbb{E}\big[f_{X_1 \vee Y_1}(X_{S\setminus\{1\}}) \mid X_1, Y_1\big] \leq 2 \cdot \mathbb{E}\big[f_{X_1 \vee Y_1}(Y_{S\setminus\{1\}}) \mid X_1, Y_1\big]$. Combining this with the above inequalities gives

$$\mathbb{E}[f(X_S)] \leq 2\mathbb{E}[f(Y_1)] + 2\mathbb{E}\big[f_{X_1 \vee Y_1}(Y_{S\setminus\{1\}}) \mid X_1, Y_1\big] \leq 2\mathbb{E}[f(Y_S)].$$ ◀

5 Conclusion

This paper introduces the concept of submodular norms and demonstrates their application in proving the efficiency of optimization problems beyond traditional ℓ_p objectives. We provide examples showcasing the utility of submodular norms in various scenarios. Specifically, we establish bounds on the competitive ratio of online facility location problems and the adaptivity gap of stochastic probing techniques when using symmetric norm objectives. These bounds crucially depend on the norm parameter ρ, and are approximately tight in the case of facility location. There are several natural directions for future work:

(i) *General Monotone Norms:* We have shown a logarithmic competitive ratio and adaptivity gap for online facility location and stochastic probing, respectively, when the objective is a symmetric norm or approximately a submodular norm. However, it remains open whether poly-logarithm bounds exist for either problem when the norm can be an arbitrary monotone norm.

(ii) *Symmetric Norm Stochastic Probing.* The logarithmic factor we get in our adaptivity gap bound for symmetric norm stochastic probing comes from the loss in approximating a symmetric norm by a submodular norm. However, it is not clear if such a loss is necessary. It would be interesting to determine if the true adaptivity gap is sub-logarithmic or even a constant.

(iii) *Parameter ρ.* Similar to online facility location, there are other optimization problems (e.g., online fractional set cover) which are known to have differing performance guarantees for ℓ_1 and ℓ_∞ objectives. We hypothesize that for such problems with symmetric norm objectives, the parameter ρ could provide a way of interpolating between ℓ_1 and ℓ_∞.

References

1. Alexandr Andoni, Jaroslaw Blasiok, and Arnold Filtser. Communication complexity of inner product in symmetric normed spaces. In *ITCS*, volume 251, pages 4:1–4:22, 2023.
2. Alexandr Andoni, Chengyu Lin, Ying Sheng, Peilin Zhong, and Ruiqi Zhong. Subspace embedding and linear regression with orlicz norm. In *ICML*, volume 80 of *Proceedings of Machine Learning Research*, pages 224–233. PMLR, 2018.
3. Alexandr Andoni, Assaf Naor, Aleksandar Nikolov, Ilya P. Razenshteyn, and Erik Waingarten. Hölder homeomorphisms and approximate nearest neighbors. In *FOCS*, pages 159–169. IEEE Computer Society, 2018.
4. Alexandr Andoni, Huy L. Nguyen, Aleksandar Nikolov, Ilya P. Razenshteyn, and Erik Waingarten. Approximate near neighbors for general symmetric norms. In *STOC*, pages 902–913. ACM, 2017.

5 Arash Asadpour and Hamid Nazerzadeh. Maximizing stochastic monotone submodular functions. *Manag. Sci.*, 62(8):2374–2391, 2016.

6 Brian Axelrod, Yang P. Liu, and Aaron Sidford. Near-optimal approximate discrete and continuous submodular function minimization. In *SODA*, pages 837–853. SIAM, 2020.

7 Yossi Azar, Niv Buchbinder, T.-H. Hubert Chan, Shahar Chen, Ilan Reuven Cohen, Anupam Gupta, Zhiyi Huang, Ning Kang, Viswanath Nagarajan, Joseph Naor, and Debmalya Panigrahi. Online algorithms for covering and packing problems with convex objectives. In *FOCS*, pages 148–157. IEEE Computer Society, 2016.

8 Francis Bach. Submodular Functions: from Discrete to Continous Domains. *Mathematical Programming, Series A*, 2018. doi:10.1007/s10107-018-1248-6.

9 Francis R. Bach. Structured sparsity-inducing norms through submodular functions. In *Advances in Neural Information Processing Systems 23: 24th Annual Conference on Neural Information Processing Systems*, pages 118–126, 2010.

10 Francis R. Bach. Learning with submodular functions: A convex optimization perspective. *Found. Trends Mach. Learn.*, 6(2-3):145–373, 2013.

11 Rajendra Bhatia. *Matrix Analysis*, volume 169. Springer, 1997.

12 Yatao Bian, Joachim M. Buhmann, and Andreas Krause. Continuous submodular function maximization. *CoRR*, abs/2006.13474, 2020. arXiv:2006.13474.

13 Yatao An Bian, Joachim M. Buhmann, and Andreas Krause. Optimal continuous dr-submodular maximization and applications to provable mean field inference. In *ICML*, volume 97 of *Proceedings of Machine Learning Research*, pages 644–653. PMLR, 2019.

14 Domagoj Bradac, Sahil Singla, and Goran Zuzic. (Near) optimal adaptivity gaps for stochastic multi-value probing. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, APPROX/RANDOM*, pages 49:1–49:21, 2019.

15 Jaroslaw Byrka, Krzysztof Sornat, and Joachim Spoerhase. Constant-factor approximation for ordered k-median. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 620–631. ACM, 2018.

16 Deeparnab Chakrabarty and Chaitanya Swamy. Approximation algorithms for minimum norm and ordered optimization problems. In *STOC*, pages 126–137. ACM, 2019.

17 Deeparnab Chakrabarty and Chaitanya Swamy. Simpler and better algorithms for minimum-norm load balancing. In *ESA*, volume 144 of *LIPIcs*, pages 27:1–27:12. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019.

18 Shichuan Deng, Jian Li, and Yuval Rabani. Generalized unrelated machine scheduling problem. In *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023*, pages 2898–2916. SIAM, 2023.

19 Hossein Esfandiari, Amin Karbasi, and Vahab S. Mirrokni. Adaptivity in adaptive submodularity. In Mikhail Belkin and Samory Kpotufe, editors, *Conference on Learning Theory, COLT*, volume 134, pages 1823–1846. PMLR, 2021.

20 Moran Feldman and Amin Karbasi. Continuous submodular maximization: Beyond dr-submodularity. In *NeurIPS*, 2020.

21 Dimitris Fotakis. On the competitive ratio for online facility location. *Algorithmica*, 50(1):1–57, 2008.

22 Satoru Fujishige. *Submodular functions and optimization*. Elsevier, 2005.

23 Anupam Gupta, Ravishankar Krishnaswamy, and Kirk Pruhs. Online primal-dual for non-linear optimization with applications to speed scaling. In *Approximation and Online Algorithms – 10th International Workshop, WAOA*, volume 7846, pages 173–186, 2012.

24 Anupam Gupta and Viswanath Nagarajan. A stochastic probing problem with applications. In *Integer Programming and Combinatorial Optimization – 16th International Conference, IPCO*, pages 205–216, 2013.

25 Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Algorithms and adaptivity gaps for stochastic probing. In *Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1731–1747, 2016.

26 Anupam Gupta, Viswanath Nagarajan, and Sahil Singla. Adaptivity gaps for stochastic probing: Submodular and XOS functions. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1688–1702. SIAM, 2017.
27 Swati Gupta, Jai Moondra, and Mohit Singh. Socially fair and hierarchical facility location problems. In *Proceedings of Economics and Computation (EC)*, 2023.
28 Sharat Ibrahimpur. Stochastic minimum norm combinatorial optimization. *Ph.D. Thesis, University of Waterloo, http://hdl.handle.net/10012/18471*, 2022.
29 Sharat Ibrahimpur and Chaitanya Swamy. Approximation algorithms for stochastic minimum-norm combinatorial optimization. In *FOCS*, pages 966–977. IEEE, 2020.
30 Sharat Ibrahimpur and Chaitanya Swamy. Minimum-norm load balancing is (almost) as easy as minimizing makespan. In *ICALP*, volume 198, pages 81:1–81:20, 2021.
31 Sharat Ibrahimpur and Chaitanya Swamy. A simple approximation algorithm for vector scheduling and applications to stochastic min-norm load balancing. In *SOSA*, pages 247–256. SIAM, 2022.
32 Thomas Kesselheim, Marco Molinaro, and Sahil Singla. Online and bandit algorithms beyond ℓ_p norms. In *SODA*, pages 1566–1593. SIAM, 2023.
33 Thomas Kesselheim and Sahil Singla. Online learning with vector costs and bandits with knapsacks. In *Proceedings of COLT*, pages 2286–2305, 2020.
34 Jerry Li, Aleksandar Nikolov, Ilya P. Razenshteyn, and Erik Waingarten. On mean estimation for general norms with statistical queries. In *COLT*, volume 99 of *Proceedings of Machine Learning Research*, pages 2158–2172. PMLR, 2019.
35 Adam Meyerson. Online facility location. In *FOCS*, pages 426–431. IEEE Computer Society, 2001.
36 Viswanath Nagarajan and Xiangkun Shen. Online covering with sum of ℓ_q- norm objectives. In *ICALP*, volume 80 of *LIPIcs*, pages 12:1–12:12, 2017.
37 Rad Niazadeh, Tim Roughgarden, and Joshua R. Wang. Optimal algorithms for continuous non-monotone submodular and dr-submodular maximization. In *NeurIPS*, pages 9617–9627, 2018.
38 Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer Science & Business Media, 2003.
39 Zhao Song, Ruosong Wang, Lin F. Yang, Hongyang Zhang, and Peilin Zhong. Efficient symmetric norm regression via linear sketching. In *NeurIPS*, pages 828–838, 2019.
40 Qixin Zhang, Zengde Deng, Zaiyi Chen, Haoyuan Hu, and Yu Yang. Stochastic continuous submodular maximization: Boosting via non-oblivious function. In *ICML*, volume 162 of *Proceedings of Machine Learning Research*, pages 26116–26134. PMLR, 2022.

A Omitted Proofs from Section 2

A.1 Properties of Continuous Submodularity

As with submodular set functions, there are many equivalent definitions for continuous submodularity which may be helpful in different settings. These are folklore properties, but we prove them for completeness.

▶ **Lemma 28.** *Let $f : \mathbb{R}_+^d \to \mathbb{R}_+$. The following are equivalent.*

1. *f is continuously submodular.*
2. *For all $x, y, z \in \mathbb{R}_+^d$ with $\mathsf{Supp}(y) \cap \mathsf{Supp}(z) = \emptyset$, we have*

$$f(x) + f(x + y + z) \leq f(x + y) + f(x + z).$$

3. *For all $x, y \in \mathbb{R}_+^d$ with $x \leq w$, and $i \in [d]$ such that $x_i = w_i$, and $a \geq 0$, we have*

$$f(w + ae_i) - f(w) \leq f(x + ae_i) - f(x).$$

4. *For all $x \in \mathbb{R}_+^d$ and $a, b \geq 0$ and distinct $i, j \in [d]$, we have*

$$f(x) + f(x + ae_i + be_j) \leq f(x + ae_i) + f(x + be_j).$$

Proof of Lemma 28. ($1 \iff 2$) Let $x, y, z \in \mathbb{R}_+^d$ with $y \perp z$. Notice that for non-negative vectors y, z, orthogonality implies that they have disjoint support. Hence, $(x+y) \vee (x+z) = x$ and $(x+y) \wedge (x+z) = x+y+z$. Then 2 follows from the definition of continuous submodularity. Likewise, if f satisfies condition 2, then applying condition 2 with $x' := x \wedge y$, $y' := y_{\{i:y_i > x_i\}}$, and $z' := x_{\{i:x_i > y_i\}}$ gives continuous submodularity.

($2 \implies 3$) Simply take $y = w - x$ and $z = ae_i$.

($3 \implies 4$) Simply take $w = a + be_j$.

($4 \implies 2$) Let $x^{(i,j)} := x + y_{<i} + z_{<j}$. Consider the sum

$$\begin{aligned} & f(x+y) + f(x+z) - f(x) - f(x+y+z) \\ = & \sum_{i,j \in [d]} \left[f(x^{(i,j)} + y_i e_i) + f(z_i^{(i,j)} + be_j) - f(x^{(i,j)}) - f(x^{(i,j)} + y_i e_i + z_j e_j) \right]. \end{aligned}$$

By condition 4, every term in the RHS sum is non-negative, so the LHS is non-negative as well. ◀

A commonly studied variant of continuous submodularity is DR-submodularity [13, 20, 37]: a function $f : \mathbb{R}_+^d \to \mathbb{R}_+$ is *DR-submodular* if it satisfies the stronger condition that for all $x, w \in \mathbb{R}_+^d$ with $x \leq w$, $i \in [d]$, and $a \geq 0$, we have $f(w + ae_i) - f(w) \leq f(x + ae_i) - f(x)$. In other words, f satisfies condition 3 of Lemma 28 even where $w_i \neq x_i$. However, the only DR-submodular norm is the ℓ_1-norm.

▶ **Lemma 29.** *Any DR-submodular norm is equivalent to ℓ_1 up to rescaling the coordinates.*

Proof. Suppose $\|\cdot\| : \mathbb{R}_+^d \to \mathbb{R}$ is a DR-submodular norm with $\|e_i\| = 1$ for each $i \in [d]$. Clearly, $\|x\| \leq \sum x_i = \|x\|_1$ by triangle inequality. Suppose that $\|x\| < \|x\|_1$ for some $x \in \mathbb{R}_+^d$. Then for some $i \in [d]$, we have $\|x_{\leq i}\| - \|x_{\leq i-1}\| \leq x_i - \varepsilon$, where $\varepsilon > 0$. By DR submodularity, this means $\|x_{\leq i-1} + kx_i e_i\| \leq k(x_i - \varepsilon)$ for all $k \in \mathbb{N}$. However, with the continuity of norms, this gives

$$x_i - \varepsilon \geq \lim_{k \to \infty} \frac{\|x_{\leq i-1} + kx_i e_i\|}{k} = \lim_{k \to \infty} \left\| \frac{1}{k} \cdot x_{\leq i-1} + x_i e_i \right\| = \|x_i e_i\| = x_i,$$

which is a contradiction, so we have $\|x\| = \|x\|_1$. ◀

B Omitted Proofs from Section 3

B.1 Uniform Costs

▷ Claim 22. Let $A \subseteq [n]$ be a fixed set of indices, and let $S \subseteq A$ be the subset of indices that arrive before the first facility is constructed at any step in A. Then $\mathbb{E}[\sum_{i \in S} \mathsf{cost}(i)] \leq 2f$.

Proof. Consider the following game: For $i \in A$, a player is shown δ_i, and has to option to pay a cost of δ_i to play a lottery, which has a $\frac{\delta_i}{f}$ chance of giving reward f. Since the expected reward of playing the lottery is exactly the cost, the player is indifferent to playing at each step. This means any strategy for the player has zero expected reward. In particular, the strategy of playing the lottery only until the first win has an expected reward of 0.

Let R be the total lottery winnings of this strategy and C be the total cost of playing. We have that $\mathbb{E}[R - C] = 0$, and since at most one lottery is won, $\mathbb{E}[R] \leq f$. Thus, $\mathbb{E}[R + C] = 2\mathbb{E}[R] \leq 2f$. But $R + C$ has exactly the distribution of $\sum_{i \in S} \mathsf{cost}(i)$, which gives the desired result. ◁

B.2 Non-uniform Costs

We now introduce the notation needed to prove that Algorithm 1 shows Theorem 8. Let us recall that $f : X \to \mathbb{R}_+$ is the cost function of opening a facility. First, we assume without loss of generality that $f(x)$ is a power of 2 for each $x \in X$.[3] Let $f^{(1)} \leq \cdots \leq f^{(m)}$ be the distinct cost levels, so $2f^{(i)} \leq f^{(i+1)}$. Additionally, let $f^{(0)} := 0$ for completeness. As before, let F_i denote the set of facilities that have been opened after the arrival of x_i, before the arrival of x_{i+1}.

▶ **Definition 30.** *For each step $i \in [n]$ and cost level $j \in \{0, \ldots, m\}$, let us define*

1. $W_i^{(j)} := F_{i-1} \cup \{x \in X : f(x) \leq f^{(j)}\}$ *to be the set of locations which are open or have an opening cost at most $f^{(j)}$;*
2. $\hat{d}_i^{(j)} := \min\{d(x_i, W_i^{(j)}), \tau_i\}$ *to be the capped value of $d_i^{(j)}$ (where cap τ_i is defined in Item 6);*
3. $\hat{d}_i := \hat{d}_i^{(0)} = \min\{d(x_i, F_{i-1}), \tau_i\}$ *for simplicity.*
4. $\delta_i^{(j)} := \frac{\hat{d}_i^{(j)}}{\hat{d}_i^{(0)}} \left(\|\hat{\boldsymbol{d}}_{\leq i}^{(0)}\| - \|\hat{\boldsymbol{d}}_{\leq i-1}^{(0)}\| \right)$ *to be the fraction of marginal increase in assignment cost we attribute to cost levels $\leq j$;*
5. $p_i^{(j)} := \frac{\delta_i^{(j-1)} - \delta_i^{(j)}}{f^{(j)}}$ *for $j \geq 1$ to be the assigned probability of opening a facility in $W_i^{(j)}$, and $p_i^{(0)} := 1 - \sum_{j=1}^m p_i^{(j)}$;*
6. $\tau_i := \arg\max\{\tau \in \mathbb{R}_{\geq 0} \cup \{+\infty\} \mid \sum_{j=1}^m p_i^{(j)} \leq 1\}$ *to be the cap value, i.e., the largest nonnegative cap such that $\sum_{j=1}^m p_i^{(j)} \leq 1$. This exists as each $p_i^{(j)}$ is monotone decreasing in τ_i.*

Given the above definitions, we notice that since $d_i^{(j)}$ is decreasing in j, this means for some j we have

$$0 = \hat{d}_i^{(m)} \leq \hat{d}_i^{(m-1)} \leq \cdots \leq \hat{d}_i^{(j+1)} \leq \tau_i = \hat{d}_i^{(j)} = \cdots = \hat{d}_i^{(0)}.$$

We shall prove the following theorem as discussed in Section 3.2.

▶ **Theorem 8.** *For online facility location problem with a submodular norm $\|\cdot\|$, there exists a randomized online algorithm that obtains cost at most $O(\log \rho) \cdot \sum_{z \in F^*} f(z) + O(1) \cdot \|\boldsymbol{d}^*\|$, where F^* and $\boldsymbol{d}^*$ are the set of facilities and vector of assignment distances respectively given by the offline optimum algorithm and $\rho := \frac{\|(1,1,\ldots,1)\|}{\min_i \|e_i\|} \leq n \cdot \frac{\max_i \|e_i\|}{\min_i \|e_i\|}$.*

To prove this theorem, we will separately the so-called short distance demands SD_k^ℓ and long distance demands LD_k^ℓ in each ring C_k^ℓ. Formally, we define

$$\mathsf{LD}_k^\ell := \{i \in C_k^\ell : \hat{d}_i^{(0)} > (\lambda+1)2^\ell r\}, \qquad \mathsf{LD} := \bigcup_{k=1}^K \bigcup_{\ell=0}^L \mathsf{LD}_k^\ell,$$

$$\mathsf{SD}_k^\ell := \{i \in C_k^\ell : \hat{d}_i^{(0)} \leq (\lambda+1)2^\ell r\}, \qquad \mathsf{SD} := \bigcup_{k=1}^K \bigcup_{\ell=0}^L \mathsf{SD}_k^\ell.$$

[3] That is, by rounding costs down to powers of 2, we lose only a factor of 2.

B.2.1 Short-distance stage

▶ **Lemma 25** (Short-Distance Stage). *The expected cost incurred by the algorithm in the short-distance stage is*

$$\mathbb{E}\left[ALG_{SD}\right] = \mathbb{E}\Big[\sum_{i\in SD} cost(i)\Big] \leq 36\|\boldsymbol{d}^*\|.$$

Proof. Let us fix a set C_k^ℓ. If $\ell = 0$, then for all $i \in \mathsf{SD}_k^\ell$, we have that $\hat{d}_i \leq (\lambda+1)r$. If $\ell > 0$, we still have that $\hat{d}_i \leq (\lambda+1)2^\ell r \leq 2(\lambda+1)d_i^*$. Summing up overall demands arriving in the short distance stage we have,

$$\begin{aligned}
\mathbb{E}\left[\mathsf{ALG}_{\mathsf{SD}}\right] &\leq \sum_{i\in\mathsf{SD}} \mathbb{E}[\mathsf{cost}(i)] \leq \sum_{i\in\mathsf{SD}} 2\delta_i^{(0)} = \sum_{i\in\mathsf{SD}:d_i^*\leq r} 2\delta_i^{(0)} + \sum_{i\in\mathsf{SD}:d_i^*>r} 2\delta_i^{(0)} \\
&\leq 2\cdot\|(\hat{d}_i)_{i\in\mathsf{SD}:d_i^*\leq r}\| + 2\cdot\|(\hat{d}_i)_{i\in\mathsf{SD}:d_i^*>r}\| \\
&\leq 2(\lambda+1)\cdot r\cdot\|(1\ldots1)\| + 4(\lambda+1)\cdot\|\boldsymbol{d}^*\| \\
&\leq 6(\lambda+1)\cdot\|\boldsymbol{d}^*\|.
\end{aligned}$$

Here, the second inequality comes from the fact that we need to account for the facility opening cost as well as the connection cost. Moreover, the third inequality holds by norm submodularity, the fourth by what was argued earlier on distances, and the last by definition of r. The lemma then follows from choosing $\lambda = 5$, which is needed for the proof of Lemma 26. ◀

B.2.2 Long-distance stage

▶ **Lemma 26** (Long-Distance Stage). *The expected cost incurred in the long-distance stage is*

$$\mathbb{E}\Big[ALG_{LD}\Big] \leq 48(\log\rho+1)\cdot\sum_{k\in[K]} f(c_k^*).$$

Proof. Let us fix a cluster ring C_k^ℓ, and let j_k^* be defined such that $f(c_k^*) = f^{(j_k^*)}$. Denote by $\gamma_i^{(j)} := d(c_k^*, W_i^{(j)})$, the distance at step i between the cluster center and a facility whose opening cost is at most $f^{(j)}$. We denote by $\mathcal{E}_\ell^{(j)}$ the event that a facility is opened within a $\gamma_0^{(j)} + 2^{\ell+1}r$ distance from optimal center c_k^*. It is easy to see that such an event occurs whenever the algorithm constructs a facility of cost $f^{(j)}$ or higher for a demand in C_k^ℓ. We now analyze the expected cost accumulated by the algorithm before, and after $\mathcal{E}_\ell^{(j)}$ has occurred. We denote by $t_\ell^{(j)}$ the time of event $\mathcal{E}_\ell^{(j)}$ occurrence.

Before $\mathcal{E}_\ell^{(j)}$ has occurred, we have $\sum_{i\leq t_\ell^{(j)}} \mathbb{E}[\mathsf{cost}^{(j)}(i)] \leq 2f^{(j)}$, by the same reasoning as Claim 22. Hence, we have that the total cost all levels $j \leq j_k^*$ before event $\mathcal{E}_\ell^{(j)}$ is

$$\sum_{j\leq j_k^*}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ i\leq t_\ell^{(j)}}} \mathbb{E}[\mathsf{cost}^{(j)}(i)] \leq 2\sum_{j\leq j_k^*} f^{(j)} \leq 4f^{(j_k^*)}.$$

We seek to demonstrate that these costs make up a constant fraction of all costs during the long-distance stage. Notice that by re-indexing, we can write

$$\sum_{j\leq j_k^*}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ i<t_\ell^{(j)}}} \mathbb{E}[\mathsf{cost}^{(j)}(i)] = \sum_{j=0}^{j_k^*-1}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ t_\ell^{(j)}<i\leq t_\ell^{(j+1)}}}\sum_{j'=j+1}^{j_k^*} \mathbb{E}[\mathsf{cost}^{(j')}(i)] = \mathbb{E}\sum_{j=0}^{j_k^*-1}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ t_\ell^{(j)}<i\leq t_\ell^{(j+1)}}} 2(\delta_i^{(j)} - \delta_i^{(j_k^*)}),$$

i.e., these are also the costs that occur in the range $\{j+1,\ldots,j_k^*\}$, during each period between event $\mathcal{E}_\ell^{(j)}$ and $\mathcal{E}_\ell^{(j+1)}$. In particular, we will show for each term in the sum, $(\delta_i^{(j)}-\delta_i^{(j_k^*)})/\delta_i^{(0)} \geq \lambda-4/\lambda+1$, so these costs comprise a constant fraction of the total expected cost $2\delta_i^{(0)}$ at each step i.

We start with the simple observation that

$$\frac{\delta_i^{(j)}-\delta_i^{(j_k^*)}}{\delta_i^{(0)}} = \frac{\hat{d}_i^{(j)}-\hat{d}_i^{(j_k^*)}}{\hat{d}_i^{(0)}} \geq \frac{d(x_i,W_i^{(j)})}{d(x_i,F_{i-1})} - \frac{d(x_i,c_k^*)}{\hat{d}_i^{(0)}},$$

by definition and subadditivity. We now proceed with bounding each term. Since $\mathcal{E}_\ell^{(j)}$ has occurred, but we are still in the long distance stage, we have

$$(\lambda+1)2^\ell r < \hat{d}_i^{(0)} \leq d(x_i,F_{i-1}) \leq d(x_i,c_k^*)+d(c_k^*,F_{i-1}) \leq \gamma_i^{(j)}+3\cdot 2^\ell r.$$

This implies both $d(x_i,F_{i-1}) \leq \gamma_i^{(j)}+3\cdot 2^\ell r$ and $\gamma_i^{(j)} \geq (\lambda-2)2^\ell r$. Additionally, we have $d(x_i,c_k^*) \leq 2^\ell r$, and by triangle inequality, we have that $\gamma_i^{(j)} \leq d(x_i,W_i^{(j)})+2^\ell r$.

Altogether, we get

$$\frac{d(x_i,W_i^{(j)})}{d(x_i,F_{i-1})} - \frac{d(x_i,c_k^*)}{\hat{d}_i^{(0)}} \geq \frac{\gamma_i^{(j)}-2^\ell r}{\gamma_i^{(j)}+3\cdot 2^{\ell+1}r} - \frac{2^\ell r}{(\lambda+1)2^\ell r} \geq \frac{\lambda-4}{\lambda+1},$$

as desired. Thus, the total cost of points in LD_k^ℓ is bounded as follows:

$$\begin{aligned}\sum_{i\in\mathsf{LD}_k^\ell}\mathbb{E}[\mathsf{cost}(i)] &= \sum_{j=0}^{j_k^*-1}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ t_\ell^{(j)}<i\leq t_\ell^{(j+1)}}}\mathbb{E}[\mathsf{cost}(i)] = \mathbb{E}\sum_{j=0}^{j_k^*-1}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ t_\ell^{(j)}<i\leq t_\ell^{(j+1)}}} 2\delta_i^{(0)}\\ &\leq 2\frac{\lambda+1}{\lambda-4}\cdot\mathbb{E}\sum_{j=0}^{j_k^*-1}\sum_{\substack{i\in\mathsf{LD}_k^\ell\\ t_\ell^{(j)}<i\leq t_\ell^{(j+1)}}}(\delta_i^{(j)}-\delta_i^{(j_k^*)})\\ &\leq 8\frac{\lambda+1}{\lambda-4}f^{(j_k^*)}.\end{aligned}$$

We now sum across all concentric rings across all K optimal clusters, to obtain that

$$\sum_{i\in\mathsf{LD}}\mathbb{E}[\mathsf{cost}(i)] = \sum_{k=1}^{K}\sum_{\ell=0}^{L}\sum_{i\in\mathsf{LD}_k^\ell}\mathbb{E}[\mathsf{cost}(i)] \leq 8L\left(\frac{\lambda+1}{\lambda-4}\right)\cdot\sum_{k\in[K]}f^{(j_k^*)},$$

and the claim follows from choosing $\lambda=5$. ◀

B.3 Lower Bound

▶ **Theorem 31.** *For any monotone norm $\|\cdot\|$, there exists a uniform-cost OFL problem with norm $\|\cdot\|$ such that any online algorithm only achieves $\Omega\left(\frac{\log\sigma}{\log\log\sigma}\right)$ competitive ratio, where $\sigma=\frac{\|\mathbf{1}_{\leq n}\|}{\max_i\|e_i\|}$. Notice that for symmetric norms, $\sigma=\rho$.*

Proof. We may assume $\max_i \|e_i\| = 1$ (otherwise we rescale costs), so $\|\mathbf{1}_{\leq n}\| = \sigma$. Let k be the largest integer such that $k^k \leq \sigma$, so we have $k = \Theta\left(\frac{\log \sigma}{\log \log \sigma}\right)$. Assume $k \geq 2$.

Now, let $G = (V, E)$ be a complete N-ary tree with height k, where N is sufficiently large (intuitively, think of N as infinite). For $j = 0, \ldots k-1$, each downwards edge from a node at depth level j will have length k^{-j}. We also define the facility opening cost to be k.

Let $0 \leq m_0 \leq m_1 \leq \cdots \leq m_k$ be defined such that m_j is the least positive integer with $\|\mathbf{1}_{\leq m_j}\| \geq k^j$. Notice that this implies $\|\mathbf{1}_{\leq m_j}\| \leq \|\mathbf{1}_{\leq m_j - 1}\| + \|e_{m_j}\| < k^j + 1$.

Our adversary will supply the demand locations as follows. First, they will choose a random path $v_0 v_1 \ldots v_k$ from the root v_0 to a leaf v_k. Then, for $j = 0, \ldots, k$, the adversary supply v_j as a demand repeated $m_j - m_{j-1}$ times (m_0 times for $j = 0$).

In the offline setting, one may simply place a single facility at v_k and assign all demands to it. This gives

$$\mathsf{OPT} = k + \|\boldsymbol{d}^*\| \leq k + \sum_{j=0}^{k-1} k^{-j} \|\mathbf{1}_{\leq m_j}\| \leq k + \sum_{j=0}^{k-1} k^{-j}(k^j + 1) = O(k).$$

In the online setting, we will show that no algorithm can achieve an expected cost of less than $\Omega(k^2)$.

Notice that any online algorithm, upon receiving a demand at v_j, should only consider the options of allocating the demand or constructing a facility at v_j. Constructing a facility anywhere else is strictly disadvantageous, as there is a negligible probability (by choice of N) that the chosen location is in the subtree rooted at v_{j+1}. Thus, after the algorithm is complete, it will have constructed a set of facilities $F \subseteq \{v_0, \ldots, v_k\}$, and each demand will be allocated to the most recently constructed facility above it. Let $\boldsymbol{d}$ be the vector of allocation distances.

If there is some $j \geq 2$ such that $v_j, v_{j-1} \notin F$, then notice that every demand at v_j will have allocation distance at least k^{-j+2}. Thus, we have

$$\|\boldsymbol{d}\| \geq k^{-j+2} \|\mathbf{1}_{\leq m_j} - \mathbf{1}_{\leq m_{j-1}}\| \geq k^{-j+2}(\|\mathbf{1}_{\leq m_j}\| - \|\mathbf{1}_{\leq m_{j-1}}\|) \geq k^2 - k - k^{-j+2} = \Omega(k^2)$$

However, if no such j exists, then $|F| \geq k/2$, so construction costs are at least $k^2/2 = \Omega(k^2)$. ◀

▶ **Corollary 32.** *In the case of a symmetric norm $\|\cdot\|$, our lower bound becomes $\Omega\left(\frac{\log \rho}{\log \log \rho}\right)$ as $\rho = \sigma = \frac{\|\mathbf{1}_{\leq n}\|}{\|e_1\|}$.*