Approximating Submodular Functions Everywhere

Michel X. Goemans* Nicholas J. A. Harvey† Satoru Iwata‡ Vahab Mirrokni§

Abstract

Submodular functions are a key concept in combinatorial optimization. One can efficiently solve many optimization problems involving a submodular function, such as computing its minimum value, or approximating its maximum value.

In this paper we consider the problem of approximating a submodular function *everywhere*, i.e., approximating its value at every point of the domain. Given oracle access to a function f on a ground set of size n, the goal is to design an algorithm which performs $\mathrm{poly}(n)$ queries to the oracle, then constructs an oracle for a function $\hat{f}$ such that, for *every* set S, $\hat{f}(S)$ approximates $f(S)$ to within a factor α.

We present two algorithms for solving this problem. The first algorithm assumes that f is a rank function of a matroid and achieves approximation factor $\alpha = \sqrt{n}+\epsilon$, for any $\epsilon > 0$. The second algorithm assumes that f is a non-negative, monotone, submodular function and achieves approximation factor $\alpha = O(\sqrt{n}\log n)$. The main technique is to compute an approximately-maximum volume ellipsoid inscribed in a symmetrized polymatroid. The analysis involves various properties of submodular functions and polymatroids.

The approximation factors achieved by our algorithms are optimal up to logarithmic factors. Indeed, we show that no algorithm can achieve a factor better than $\Omega(\sqrt{n}/\log n)$, even for matroid rank functions.

1 Introduction

Let $f : 2^{[n]} \to \mathbb{R}_+$ be a function where $[n] = \{1, 2, \cdots, n\}$. The function f is called *submodular* if

$$f(S) + f(T) \;\geq\; f(S \cup T) + f(S \cap T) \qquad \forall S, T \subseteq [n]. \tag{1}$$

Additionally, f is called *monotone* if $f(Y) \leq f(Z)$ whenever $Y \subseteq Z$. An equivalent definition of submodularity is the property of *decreasing marginal values*:

$$f(Y \cup \{x\}) - f(Y) \;\geq\; f(Z \cup \{x\}) - f(Z) \qquad \forall Y \subseteq Z \subseteq [n] \text{ and } x \in [n] \setminus Z. \tag{2}$$

Eq. (2) can be deduced from Eq. (1) by substituting $S = Y \cup \{x\}$ and $T = Z$. The reverse implication also holds; see, e.g., Schrijver [37, Theorem 44.1]. Throughout this paper we will assume that each submodular function is presented by a *value oracle*; i.e., for any set S, an algorithm can query an oracle to find its value $f(S)$.

1.1 Background

Submodular functions are a key concept in operations research and combinatorial optimization, since many combinatorial optimization problems can be naturally formulated in terms of submodular functions. There are several books devoted to this subject [12, 35, 34]. One explanation for the usefulness of submodular functions is that they can be viewed as a discrete analogue to convex functions; for more on this connection, see Lovász [32] and Murota [34].

*MIT Department of Mathematics. goemans@math.mit.edu. Supported by NSF contracts CCF-0515221 and CCF-0829878 and by ONR grant N00014-05-1-0148.

†University of Waterloo, Department of Combinatorics and Optimization. harvey@math.uwaterloo.ca. Portions of this work were performed while at Microsoft Research, New England, and at the MIT Computer Science and Artificial Intelligence Laboratory. Portions of this work were supported by an NSERC Discovery Grant.

‡RIMS, Kyoto University, Kyoto 606-8502, Japan. iwata@kurims.kyoto-u.ac.jp. Supported by the Kayamori Foundation of Information Science Advancement.

§Google Research, New York, NY. mirrokni@gmail.com.

Both minimizing and maximizing submodular functions, possibly under some extra constraints, have been considered extensively in the literature. Minimizing a submodular function can be solved exactly with only $\text{poly}(n)$ oracle calls, either by the ellipsoid algorithm [17] or through combinatorial algorithms that have been obtained in the last decade [20, 38]. Unlike submodular function minimization, the problem of maximizing a submodular function is an NP-hard problem since it generalizes many NP-hard problems, such as the maximum cut problem. In many settings, constant-factor approximation algorithms have been developed for this problem. For example, there is a $\frac{2}{5}$-approximation algorithm for maximizing any non-negative, non-monotone submodular function [11], and a $(1 - 1/e)$-approximation algorithm for maximizing a monotone submodular function subject to a cardinality constraint [36], or an arbitrary matroid constraint [45]. Approximation algorithms for submodular analogues of several other well-known optimization problems have been studied, e.g., [46, 43].

Submodular functions have been of recent interest due to their applications in combinatorial auctions, particularly the submodular welfare problem [30, 27, 8]. The goal of this problem is to partition a set of items among a set of players in order to maximize their total utility. In this context, it is natural to assume that the players' utility functions are submodular, as this captures a realistic notion of diminishing returns. Under this submodularity assumption, efficient approximation algorithms have recently been developed for this problem [8, 45].

1.2 Contributions

The extensive literature on submodular functions motivates us to investigate other fundamental questions concerning their structure. How much "information" is contained in a submodular function? How much of that information can be obtained in just a few value oracle queries? Can an auctioneer efficiently estimate a player's utility function if it is submodular? To address these questions, we consider the problem of approximating a submodular function f everywhere while performing only a polynomial number of queries. More precisely, the problem we study is:

Problem 1 *Given oracle access to a non-negative, monotone, submodular function $f : 2^{[n]} \to \mathbb{R}$, design an algorithm that performs $\text{poly}(n)$ queries to f, then constructs an oracle for a function $\hat{f}$ (not necessarily submodular) which is an approximation of f, in the sense that $\hat{f}(S) \le f(S) \le \alpha \cdot \hat{f}(S)$ for all $S \subseteq [n]$. For what values of α (possibly a function of n) is this possible?*

For some submodular functions this problem can be solved exactly, i.e., with $\alpha = 1$. As an example, for undirected graph cut functions, it is easy to completely reconstruct the graph while using only $O(n^2)$ queries. For more general submodular functions, we prove the following results.

- When f is the rank function of a matroid, we give an algorithm that computes a function $\hat{f}$ which achieves approximation factor $\alpha = \sqrt{n} + \epsilon$, for any $\epsilon > 0$.
- When f is an arbitrary non-negative, monotone, submodular function, we give an algorithm that computes a function $\hat{f}$ giving an approximation factor $\alpha = O(\sqrt{n} \log n)$.
- On the other hand, we show that any algorithm which performs only $\text{poly}(n)$ queries must satisfy $\alpha = \Omega(\sqrt{n}/\log n)$, even if f is known to be the rank function of a matroid. If f is not assumed to be monotone, the lower bound improves slightly to $\alpha = \Omega(\sqrt{n/\log n})$.

For both of the algorithms mentioned above, the computed function $\hat{f}$ is actually submodular, and it takes the particularly simple form of a *root-linear* function. Formally,

$$\hat{f}(S) \;=\; \sqrt{\sum_{i \in S} c_i}$$

for some $c \in \mathbb{R}^n_+$. This is a useful structural result because many optimization problems are easier to solve for root-linear functions than for general submodular functions. This suggests a general technique for solving some submodular optimization problems: use our algorithm to approximate the submodular function by a root-linear function, then solve the optimization problem on the root-linear function. Some applications of this technique are described in Section 7.

1.3 Related work

Our lower bounds stated above were previously described in an unpublished manuscript [15]. This manuscript also gave a *non-adaptive* algorithm that solves Problem 1 with $\alpha = n/(c \log n)$ for any constant c; furthermore, this is optimal (amongst non-adaptive algorithms).

A subsequent paper of Svitkina and Fleischer [43] proposed several new optimization problems on submodular functions, such as Submodular Load Balancing, Submodular Sparsest Cut, Submodular Knapsack, etc. They give algorithms for these problems with approximation factor $\tilde{O}(\sqrt{n})$. Motivated by the manuscript [15], Svitkina and Fleischer also studied Problem 1 and gave a randomized algorithm with approximation factor $\alpha = 2\sqrt{n}$, but only for a very restricted class of submodular functions. Additionally, they adjusted the parameters of our lower bound construction, yielding an improved $\Omega(\sqrt{n/\log n})$ lower bound for Problem 1 when f is monotone and submodular. They also use variants of our lower bound construction to obtain nearly-optimal lower bounds for all of the problems that they studied.

Many more submodular optimization problems were proposed in later work [14, 22]. Several portions of this work build on the techniques in this paper. For example, the lower bounds in [22] for Submodular Edge Cover and Submodular Cost Set Cover, and the lower bounds in [14] for Submodular Vertex Cover, Submodular Shortest Path, Submodular Perfect Matching and Submodular Minimum Spanning Tree are all variants on our lower bound for Problem 1. Additionally, the algorithm for the Submodular Shortest Path problem [14] uses our algorithm for Problem 1 as a subroutine.

Subsequent work [23, 24] studied the Edge-Submodular (s,t)-cut problem, which they also call the Cooperative (s,t)-cut problem. One of their algorithms for this problem uses our algorithm for Problem 1 as a subroutine. Furthermore, their lower bound for this problem is again a variant on our lower bound for Problem 1.

1.4 Algorithmic Approach

A high-level overview of our algorithmic approach is as follows. Suppose there exists a centrally symmetric convex body $K \subset \mathbb{R}^n$ such that $f(S) = \max \{ \chi(S)^\mathsf{T} x : x \in K \}$, where $\chi(S)$ is the characteristic vector of S. We shall see that such a body K can easily be obtained from the submodular polyhedron defined by f. John's theorem [25, p203] states that there exists an ellipsoid E centered at the origin such that

$$E \;\subseteq\; K \;\subseteq\; \sqrt{n}E. \tag{3}$$

Then clearly

$$\max \{ \chi(S)^\mathsf{T} x : x \in E \} \;\leq\; \max \{ \chi(S)^\mathsf{T} x : x \in K \} \;\leq\; \max \{ \chi(S)^\mathsf{T} x : x \in \sqrt{n}E \}. \tag{4}$$

Defining $\hat{f}(S) = \max \{ \chi(S)^\mathsf{T} x : x \in E \}$, we have

$$\hat{f}(S) \;\leq\; f(S) \;\leq\; \sqrt{n} \cdot \hat{f}(S),$$

so f is approximated everywhere by $\hat{f}$ to within a factor $\sqrt{n}$. The main task of our algorithms is to approximately compute such an ellipsoid E.

2 Mathematical Preliminaries

In this section, we state and review basic facts from convex geometry, focusing on properties of ellipsoids. For a more detailed discussion of these topics, we refer the reader to standard references [5, 18, 17, 33, 39, 47].

Ellipsoids. All matrices that we discuss are real, symmetric, and have size $n \times n$. If a matrix A is positive semi-definite we write $A \succcurlyeq 0$, and if it is positive definite we write $A \succ 0$. Let $A \succ 0$ and let $A^{1/2}$ be its (unique) symmetric, positive definite square root, meaning $A = A^{1/2} A^{1/2}$. We define the *ellipsoidal norm* $\|\cdot\|_A$ in $\mathbb{R}^n$ by $\|x\|_A = \sqrt{x^\mathsf{T} A x}$. Let B_n denote the closed, Euclidean unit ball $\{ x \in \mathbb{R}^n : \|x\| \leq 1 \}$, and let V_n denote its volume. Given $A \succ 0$, let $E(A)$ denote the ellipsoid centered at the origin defined by A:

$$E(A) \;=\; \{ x \in \mathbb{R}^n : x^\mathsf{T} A x \leq 1 \} \;=\; \{ x : \|x\|_A \leq 1 \}.$$

Equivalently, we can obtain $E(A)$ by applying the linear map $x \mapsto A^{-1/2}x$ to B_n. Consequently, the volume of $E(A)$ is $V_n / \det(A^{1/2})$. Given any non-zero $c \in \mathbb{R}^n$, we have that

$$\begin{aligned} \max\{\, c^\mathsf{T} x : x \in E(A) \,\} &= \max\{\, c^\mathsf{T} A^{-1/2} x : x \in B_n \,\} \\ &= c^\mathsf{T} A^{-1/2} \Big(\frac{A^{-1/2} c}{\|A^{-1/2} c\|} \Big) \\ &= \sqrt{c^\mathsf{T} A^{-1} c} \\ &= \|c\|_{A^{-1}}. \end{aligned}$$

Polarity. A set $K \subseteq \mathbb{R}^n$ is called *centrally symmetric* if $x \in K \Leftrightarrow -x \in K$. A set $K \subset \mathbb{R}^n$ is called a *convex body* if K is convex, compact, and has non-empty interior. The *polar* of a set $K \subseteq \mathbb{R}^n$ is defined to be

$$K^* \; := \; \big\{\, c \in \mathbb{R}^n : c^\mathsf{T} x \leq 1 \;\; \forall x \in K \,\big\}.$$

Some basic facts about polarity are listed below.

(P1) For any set K, the polar K^* is closed, convex and contains 0.

(P2) If K is centrally symmetric, then the polar K^* is centrally symmetric.

(P3) The polar of B_n is B_n itself. Moreover, B_n is the unique set which has this property.

(P4) Let K and L be arbitrary sets. Then $(K \cup L)^* = K^* \cap L^*$.

(P5) Let K be a convex body. Then K^* is bounded if and only if 0 lies in the interior of K.

(P6) For any set K, we have $K^{**} = \operatorname{cl}\big(\operatorname{conv}(K \cup \{0\})\big)$. Consequently $K^* = K^{***} = \operatorname{cl}(\operatorname{conv}(K \cup \{0\}))^*$. Furthermore, if K is a convex body and $0 \in K$ then $K^{**} = K$.

(P7) For any set $K \subset \mathbb{R}^n$ and invertible linear map $L : \mathbb{R}^n \to \mathbb{R}^n$, we have $(L(K))^* = L^{-\mathsf{T}}(K^*)$. In particular,

$$E(A)^* \; = \; \big(A^{-1/2}(B_n)\big)^* \; = \; A^{1/2}(B_n^*) \; = \; E(A^{-1}).$$

(P8) Let K and L be convex bodies that both contain 0. Then $L \subseteq K \Leftrightarrow K^* \subseteq L^*$.

(P9) Let K and L be convex bodies that both contain 0. Then $(K \cap L)^* = \operatorname{cl}\big(\operatorname{conv}(K^* \cup L^*)\big)$.

(P10) Let P be a polytope that contains 0 in its interior. Then P^* is also a polytope. Moreover, c is a vertex of P if and only if $c^\mathsf{T} x \leq 1$ is a facet of P^*.

Minimum volume circumscribed ellipsoid. Let K be a convex body. It is known that there is a unique ellipsoid circumscribing K with minimum volume. (In this context, "circumscribing" simply means "containing".) We will refer to this ellipsoid as the *Löwner ellipsoid*, although it is also called the Löwner-John ellipsoid or Löwner-Behrend-John ellipsoid.

Suppose in addition that K is centrally symmetric. It is easy to show that the Löwner ellipsoid must be centered at the origin. We now show that the Löwner ellipsoid of K can be characterized by a convex program. An ellipsoid $E(A)$ is feasible if, for every $x \in K$,

$$\begin{aligned} & x \in E(A) \\ \iff \quad & x \in E(A^{-1})^* \\ \iff \quad & \max\big\{\, x^\mathsf{T} y : y \in E(A^{-1}) \,\big\} \leq 1 \\ \iff \quad & \|x\|_A \leq 1 \end{aligned} \tag{5}$$

We wish to minimize the volume of $E(A)$, which is proportional to $1/\det(A^{1/2})$, as observed above. Thus we arrive at the following mathematical program, which is well-known [7, §8.4.1].

$$(\text{CP1}) \qquad \begin{array}{lll} \min & -\log\det A & \\ \text{s.t.} & \|x\|_A^2 \;\leq 1 & \forall x \in K \\ & A \;\succ 0 & \end{array}$$

This is a convex program. The feasible region is clearly convex (since the set of positive definite matrices is a convex cone). The convexity of the objective function is a consequence of Lemma 1, which shows that the determinant is *strictly log-concave* over the cone of positive definite matrices.

Lemma 1 (Fan [10][1]) *Let $A, B \succ 0$, $A \neq B$, and $0 < \lambda < 1$. Then*

$$\log\det\big(\lambda A + (1-\lambda)B\big) > \lambda \log\det A \;+\; (1-\lambda)\log\det B.$$

Using the ellipsoid method [17], an approximately optimal solution to (CP1) can be efficiently computed provided we can efficiently separate over the constraints $\|x\|_A^2 \leq 1$. The separation problem amounts to maximizing the convex function $x \mapsto x^{\mathsf{T}}Ax$ over the body K. However, maximizing a convex function over a convex set is a computationally intractable problem, so (CP1) is in general difficult to solve. We remark that, if K is a polytope, the separation problem amounts to maximizing $x \mapsto x^{\mathsf{T}}Ax$ over the *vertices* of K, because the maximum of a convex function over a polytope is always attained at a vertex.

The strict log-concavity of the determinant shows that (CP1) has a unique optimum solution, since a strict convex combination of any two distinct optimum solutions would give a strictly better solution. This shows uniqueness of the minimum volume ellipsoid which is centered at the origin and contains a given centrally symmetric, convex body.

Maximum volume inscribed ellipsoid. For centrally symmetric convex bodies, polarity provides a close connection between minimum circumscribed ellipsoids and maximum inscribed ellipsoids. If K is a centrally symmetric convex body, there is a maximum volume ellipsoid inscribed in K that is centered at the origin, as was the case with circumscribed ellipsoids. (In this context, "inscribed in K" simply means "contained in K".) It follows from (P8) that $E(A)$ is a maximum volume ellipsoid inscribed in K if and only if $E(A^{-1})$ is a minimum volume ellipsoid circumscribing K^*.

On the other hand, if K is not centrally symmetric, then a maximum volume ellipsoid inscribed in K might not contain the origin, in which case its polar would be unbounded, and so certainly not an ellipsoid. Nevertheless, for any convex body there does indeed exist a unique maximum volume inscribed ellipsoid; see, e.g., [5, Theorem V.2.2]. This ellipsoid is often called the *John ellipsoid*, although this attribution is somewhat inaccurate since John [25] actually considers only circumscribed ellipsoids. However, since circumscribed and inscribed ellipsoids are interchangeable notions in the centrally symmetric case, the inaccuracy is forgivable.

For a centrally symmetric convex body K, the John ellipsoid can also be characterized by a convex program. Indeed, by our observation above, finding the John ellipsoid and finding the Löwner ellipsoid (for the polar) are equivalent problems, and the latter can be solved by (CP1). Thus, the John ellipsoid $E(A)$ is characterized by the following well-known [7, §8.4.2] mathematical program, which maximizes a concave function over a convex set.

$$\text{(CP2)} \qquad \begin{array}{rlll} \max & \log\det(A^{-1}) & & \\ \text{s.t.} & \|c\|_{A^{-1}}^2 & \leq 1 & \forall c \in K^* \\ & A^{-1} & \succ 0 & \end{array}$$

For polytopes, this can be simplified. If K is a polytope with 0 in its interior then K^* is also a polytope, by (P10). So, by our discussion of (CP1), we need only include the constraint $\|c\|_{A^{-1}}^2 \leq 1$ when c is a vertex of K^*. This condition is equivalent to $c^{\mathsf{T}}x \leq 1$ defining a facet of K, by (P10).

John's theorem. Let K be a centrally symmetric convex body and let $E(A)$ be the maximum volume ellipsoid inscribed in K. John's theorem states that K is contained in $\sqrt{n} \cdot E(A) = E(A/n)$; equivalently, $\|x\|_A \leq \sqrt{n}$ for all $x \in K$, by Eq. (5). John's theorem has an equivalent statement in terms of minimum volume ellipsoids: if $E(A)$ is the minimum volume ellipsoid circumscribing K then $\frac{1}{\sqrt{n}}E(A) \subseteq K$. This connection follows from (P8).

John's theorem plays an important role in the local theory of Banach spaces. Another equivalent statement is that the Banach-Mazur distance between any n-dimensional Banach space and the n-dimensional Hilbert space l_2^n is at most $\sqrt{n}$. A consequence is that the Banach-Mazur distance between any two n-dimensional Banach spaces is at most n; this bound is known to be tight up to a constant factor [13].

[1] Fan does not actually state the strict inequality, although his proof does show that it holds.

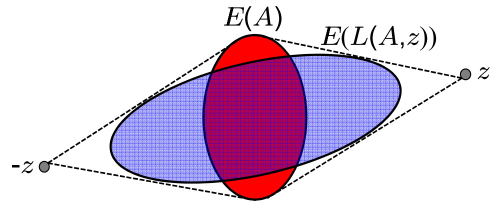

Figure 1: The red ellipsoid is $E(A)$. The point z satisfies $\|z\|_A > \sqrt{n}$. The blue ellipsoid $E(L(A,z))$ is the John ellipsoid for the convex hull of $E(A)$, z and $-z$. Its volume is larger than the volume of $E(A)$.

John's theorem can be proven in several ways. See, for example, Ball [4, Lecture 3] or Matoušek [33, §13.4]. We adopt a more algorithmic argument. Suppose there is an element $z \in K$ with $\|z\|_A > \sqrt{n}$. Then the following lemma gives an explicit construction of an ellipsoid $E(L(A,z))$ of strictly larger volume that is contained in the convex hull of $E(A)$, z and $-z$, implying that $E(L(A,z))$ is contained in K. This is illustrated in Figure 1. The new ellipsoid $E(L(A,z))$ is larger than $E(A)$ since Eq. (6) shows that the change in volume is greater than 1 when $\|z\|_A^2 = l > n$. This proves John's theorem.

Lemma 2 *For $A \succ 0$ and $z \in \mathbb{R}^n$ with $l = \|z\|_A^2 \geq n$, let*

$$L(A,z) \;=\; \frac{n}{l}\frac{l-1}{n-1}A + \frac{n}{l^2}\left(1 - \frac{l-1}{n-1}\right)Azz^{\mathsf{T}}A.$$

Then $L(A,z)$ is positive definite, the ellipsoid $E(L(A,z))$ is contained in $\operatorname{conv}(E(A) \cup \{z,-z\})$*, and its volume* $\operatorname{vol} E(L(A,z))$ *equals $\gamma_n(l) \cdot \operatorname{vol} E(A)$ where*

$$\gamma_n(l) \;=\; \sqrt{\left(\frac{l}{n}\right)^n \left(\frac{n-1}{l-1}\right)^{n-1}}. \tag{6}$$

This lemma is not new — its polar statement plays an important role in the ellipsoid method for solving linear programs. We discuss this further and give a proof of the lemma in Appendix A.1.

3 Ellipsoidal Approximations and Polymatroids

As described in Section 1.4, our main task is as follows. Given a submodular function f, we construct a convex body K which describes f, then we wish to construct an ellipsoid E satisfying Eq. (3). In Section 3.1, we describe a generic procedure for any convex body K that approximately finds such an ellipsoid E, given an oracle for approximately maximizing quadratic functions over K. Next, in Section 3.2 we describe the specific convex body K that describes the submodular function f. Finally, in Section 3.3, we show how the symmetries of K can be exploited to implement such an oracle.

3.1 Ellipsoidal Approximations

Definition 1 *Let K be a convex body. If $E(A) \subseteq K \subseteq \lambda E(A)$ then the ellipsoid $E(A)$ is called a λ-ellipsoidal approximation to K.*

The John ellipsoid is therefore a $\sqrt{n}$-ellipsoidal approximation to a centrally symmetric convex body K, and so is $1/\sqrt{n}$ times the Löwner ellipsoid. These are existential results. Algorithmically, the situation very much depends on how the convex body is given.

The simplest case is when K is a polyhedral set given explicitly as the intersection of halfspaces. In this case, the feasible region of (CP2) can be described with one inequality for each given halfspace, and so the ellipsoid method can be used to solve (CP2) to within any desired accuracy. This gives an alternate way to derive the result of Grötschel, Lovász and Schrijver [17, Theorem 4.6.5] which gives a polynomial-time algorithm to compute a $\sqrt{n+1}$-ellipsoidal approximation to a centrally symmetric convex body K given explicitly by a system of linear inequalities.

Algorithm Ellipsoidal-Approximation
Input: A centrally symmetric convex body K and an ellipsoid $E_0 \subseteq K$
Output: A $(\sqrt{n}+\epsilon)/\alpha$-ellipsoidal approximation of K
Set $j \leftarrow 0$
Repeat
  $\triangleright$ $E_j = E(A_j)$ *is an ellipsoid contained in* K
  Execute the α-approximate decision procedure on E_j and K
  If it returns a $z \in K$ with $\|z\|_{A_j} > \sqrt{n}+\epsilon$
    Set $A_{j+1} \leftarrow L(A_j, z)$, as in Lemma 2
    Set $j \leftarrow j+1$
  Else
    Return E_j
End

Figure 2: An algorithm for constructing a $(\sqrt{n}+\epsilon)/\alpha$-ellipsoidal approximation of K, given an α-approximate decision procedure for maximizing $\|x\|_A$ over K.

A much more general case is when K is itself described by a separation oracle. In this case, it is customary to assume that K is *well-bounded*, meaning that parameters $R \geq r > 0$ are given such that $B(0,r) \subseteq K \subseteq B(0,R)$; algorithms in this model can have running time that depends polynomially on $\log(R/r)$. The best known algorithmic result for this model is a polynomial-time algorithm giving only a $\sqrt{n(n+1)}$-ellipsoidal approximation [17, Theorem 4.6.3]. This result is too weak to prove our desired result on approximating submodular functions everywhere. Moreover, there is no substantially better algorithm for this general model: no algorithm, even randomized, can produce a λ-ellipsoidal approximation with λ better than $\tilde{O}(n)$ for a well-bounded, centrally symmetric convex body given by a separation oracle [40].

An improved approximation can be obtained if we assume a stronger oracle for K. Suppose K is given by an *α-approximate decision procedure* which, given $A \succ 0$ with $E(A) \subseteq K$, either returns a $x \in K$ with $\|x\|_A > \sqrt{n}+\epsilon$ or guarantees that every $x \in K$ satisfies $\|x\|_A \leq (\sqrt{n}+\epsilon)/\alpha$. Furthermore, suppose that K is well-bounded, in the sense that we are given an ellipsoid E_0 and parameter ρ such that $E_0 \subseteq K \subseteq \rho \cdot E_0$. Then we can construct a $(\sqrt{n}+\epsilon)/\alpha$-ellipsoidal approximation of K, for any $\epsilon > 0$, by the same process as our proof of John's theorem in Section 2. This algorithm is shown in Figure 2.

The correctness of this algorithm follows from the definition of the α-approximate decision procedure: when the algorithm returns, we are guaranteed that every $x \in K$ satisfies $\|x\|_{A_j} \leq (\sqrt{n}+\epsilon)/\alpha$. Since $E_j \subseteq K$, this means that E_j is an $(\sqrt{n}+\epsilon)/\alpha$-ellipsoidal approximation of K. To bound the number of iterations required, we analyze the volume of the ellipsoids E_j. When applying Lemma 2 we have $\|z\|_{A_j} > \sqrt{n}+\epsilon$, so

$$\frac{\mathrm{vol}(E_{j+1})}{\mathrm{vol}(E_j)} \;\geq\; \gamma_n((\sqrt{n}+\epsilon)^2) \;\geq\; \gamma_n(n+2\epsilon\sqrt{n}) \;\geq\; 1+\frac{2\epsilon^2}{3n^2},$$

the last inequality by Lemma 3 below. This increase in volume and the fact that $K \subseteq \rho E_0$ ensure that the number of iterations is at most $O(n^2\epsilon^{-2}\log(\rho^n)) = O(n^3\epsilon^{-2}\log\rho)$.

Lemma 3 *The function $\gamma_n(l)$ given in Lemma 2 satisfies $\gamma_n(n+x) \geq 1 + x^2/(6n^3)$ whenever $0 < x \leq n$.*

3.2 Symmetrized polymatroids

Let $f : 2^{[n]} \to \mathbb{R}$ be a monotone, submodular function with $f(\emptyset) = 0$. For any vector $x \in \mathbb{R}^n$ and $S \subseteq [n]$, let $x(S) = \sum_{i \in S} x_i$. The polytope $P_f \subseteq \mathbb{R}^n$ defined by

$$P_f \;=\; \left\{ \begin{array}{ll} x(S) \leq f(S), & \forall S \subseteq [n] \\ x \geq 0 & \end{array} \right\}$$

is called the *polymatroid*, or *submodular polyhedron*, associated with f. One important property [37, Corollary 44.3f] of this polyhedron is that

$$f(S) \;=\; \max\left\{\, \chi(S)^{\mathsf{T}} x : x \in P_f \,\right\}.$$

We will henceforth assume, without loss of generality, that P_f is full-dimensional. It is easy to see that this assumption holds iff $f(\{i\}) > 0$ for all i. However if $f(\{i\}) = 0$ then we may simply restrict to the function which ignores coordinate i, since $f(S) = f(S + i)$ for every $S \subseteq [n]$.

Now for any set $Q \subseteq \mathbb{R}^n$, let

$$S(Q) \;=\; \{\, x \in \mathbb{R}^n : |x| \in Q \,\},$$

where $|x|$ denotes component-wise absolute value. Thus $S(P_f)$ is a centrally symmetric convex body. Furthermore, it is easy to see that

$$f(S) \;=\; \max\left\{\, \chi(S)^{\mathsf{T}} x : x \in S(P_f) \,\right\}. \tag{7}$$

Suppose now that $E(A)$ is a λ-ellipsoidal approximation to $S(P_f)$. As in (4), for any $c \in \mathbb{R}^n$ we have

$$\begin{aligned} \|c\|_{A^{-1}} \;&=\; \max\{c^{\mathsf{T}} x : x \in E(A)\} \\ &\leq\; \max\{c^{\mathsf{T}} x : x \in S(P_f)\} \\ &\leq\; \lambda \max\{c^{\mathsf{T}} x : x \in E(A)\} \;=\; \lambda\, \|c\|_{A^{-1}}. \end{aligned}$$

In particular, taking $c = \chi(S)$ for any $S \subseteq [n]$, we obtain

$$\|\chi(S)\|_{A^{-1}} \;\leq\; f(S) \;\leq\; \lambda\, \|\chi(S)\|_{A^{-1}},$$

by Eq. (7). Thus the function $\hat{f}$ defined by

$$\hat{f}(S) \;=\; \|\chi(S)\|_{A^{-1}} \tag{8}$$

provides a λ-approximation to $f(S)$ everywhere. To summarize this discussion, we have shown that a λ-ellipsoidal approximation to $S(P_f)$ yields a λ-approximation to f everywhere.

3.3 Exploiting Symmetry in Ellipsoidal Approximations

In Section 3.1, we showed that a $(\sqrt{n} + \epsilon)/\alpha$-ellipsoidal approximation of a centrally symmetric convex body K can be obtained using an α-approximate decision procedure for K. In this section, we show that the symmetries of P_f can be exploited to help design such a decision procedure.

If a centrally symmetric convex body K is invariant under a linear transformation T (i.e., $T(K) = K$) then, by uniqueness, the maximum volume inscribed ellipsoid E is also invariant under T. More generally, define $\mathrm{Aut}(K)$, the automorphism group of K, to be the set of all linear operators mapping K to itself. Then the maximum volume ellipsoid E inscribed in K satisfies $T(E) = E$ for all $T \in \mathrm{Aut}(K)$. (See, e.g., Güler and Gürtina [19].) In our case, $\mathrm{Aut}(S(P_f))$ contains the subgroup of transformations T of the form $T(x) = Cx$ where C is a diagonal ± 1 matrix. We call such convex bodies *axis aligned*. This means that the maximum volume ellipsoid $E(A)$ inscribed in $S(P_f)$ is also axis aligned. This implies that A is a diagonal matrix, as proven below.

Unfortunately, if the algorithm Ellipsoidal-Approximation is given an axis-aligned body K as input, it does not necessarily output an axis-aligned ellipsoidal approximation to K. Indeed, for a diagonal matrix A, Lemma 2 does not necessarily construct an axis-aligned ellipsoid $E(L(A, z))$. However, the following proposition shows that an arbitrary ellipsoid in K can be mapped to an axis-aligned ellipsoid without decreasing its volume. (This shows that the maximum volume ellipsoid is axis aligned.)

To state the proposition, we need some notation. For a matrix A, let $\mathrm{Diag}(A)$ denote the diagonal matrix whose diagonal entries are the same as A's.

Proposition 4 *Let K be an axis-aligned convex body, and let $E(A)$ be an ellipsoid inscribed in K. Then the axis aligned ellipsoid $E(B)$ defined by the diagonal matrix $B = (\mathrm{Diag}(A^{-1}))^{-1}$ satisfies $E(B) \subseteq K$ and* $\operatorname{vol} E(B) \geq \operatorname{vol} E(A)$.

Proposition 4 shows that the Ellipsoidal-Approximation algorithm can be modified so that, if its input K is axis-aligned, then the ellipsoid E_j computed in every iteration is also axis aligned. This modification has two important consequences. First, this means that it suffices to design an α-approximate decision procedure

Algorithm Approximate-Everywhere

Input: A monotone, submodular function f with $f(\emptyset) = 0$ and $f(\{i\}) > 0$ for every $i \in [n]$

Output: A root-linear function $\hat{f}$ giving a $(\sqrt{n} + \epsilon)/\alpha$-approximation to f everywhere

Set $j \leftarrow 0$

Let D_0 be the diagonal matrix whose i^{th} diagonal entry is $(n/f(\{i\}))^2$

Repeat

  $\triangleright$ *D_j is a diagonal matrix and $E_j = E(D_j)$ is an axis aligned ellipsoid contained in $S(P_f)$*

  Execute an α-approximate decision procedure on E_j and $S(P_f)$

  If it returns a $z \in K$ with $\|z\|_{D_j} > \sqrt{n} + \epsilon$

    Set $B \leftarrow L(D_j, z)$, as in Lemma 2

    Set $D_{j+1} \leftarrow (\mathrm{Diag}(B^{-1}))^{-1}$

    Set $j \leftarrow j + 1$

  Else

    Return the function $\hat{f}$ given by $\hat{f}(S) = \sqrt{\sum_{i \in S} c_i}$ where $c_i = 1/D_{i,i}$

End

Figure 3: The algorithm for constructing a function $\hat{f}$ which is a $\sqrt{n+1}/\alpha$-approximation to f.

for $E(A)$ and K where A is assumed to be diagonal. Secondly, when applying this algorithm to $K = S(P_f)$ as in Section 3.2, the function $\hat{f}$ constructed in Eq. (8) is simply

$$\hat{f}(S) \;=\; \|\chi(S)\|_{A^{-1}} \;=\; \sqrt{\sum_{i \in S} c_i},$$

where $c_i = 1/A_{i,i}$ for every $i \in [n]$. Thus $\hat{f}$ is a root-linear function, and it is an easy exercise to verify that any such function is submodular.

In summary, algorithm Approximate-Everywhere, presented in Figure 3, gives our method for approximating submodular functions everywhere. The correctness of this algorithm follows from the preceding discussions, and from the observation that the initial ellipsoid $E_0 = E(D_0)$ satisfies $E_0 \subseteq S(P_f)$, as proven in Lemma 5. Furthermore, since $S(P_f) \subseteq n^2\, E_0$, the number of iterations is at most $O(n^3 \epsilon^{-2} \log n)$, as discussed in Section 3.1.

Lemma 5 *Let D_0 be the diagonal matrix whose i^{th} diagonal entry is $(n/f(\{i\}))^2$. Then $E(D_0) \subseteq S(P_f)$ and $S(P_f) \subseteq n^2\, E(D_0)$.*

In order to implement algorithm Approximate-Everywhere, we require an α-approximate decision procedure for maximizing $\|x\|_D$ over $S(P_f)$, where D is a positive definite diagonal matrix. By symmetry, it suffices to obtain such a procedure for P_f instead of $S(P_f)$.

Theorem 6 *Suppose we have an α-approximate decision procedure for $\max\{\, \|x\|_D : x \in P_f \,\}$, where D is diagonal and positive definite. Then* Approximate-Everywhere *outputs a $(\sqrt{n} + \epsilon)/\alpha$-approximation to f everywhere after at most $O(n^3 \epsilon^{-2} \log n)$ iterations.*

4 Matroid Rank Functions

In this section, we consider the problem of approximating a submodular function f everywhere, in the special case that f is the rank function of a matroid. By Theorem 6, it suffices to find an α-approximate decision procedure for $\max\{\, \|x\|_D : x \in P_f \,\}$, where D is a diagonal, positive definite matrix. We show how to solve this optimization problem exactly in polynomial time.

Let $M = ([n], \mathcal{I})$ be a matroid and $\mathcal{I}$ its family of independent sets. Let $f(\cdot)$ be its rank function, i.e.,

$$f(S) \;=\; \max\{\, |U| : U \subseteq S,\, U \in \mathcal{I} \,\}.$$

It is known that f is monotone and submodular. Edmonds [9] showed that the vertices of the polymatroid P_f are precisely $\{\, \chi(I) : I \in \mathcal{I} \,\}$. In particular, this implies that the vertices are all 0-1 vectors.

Clearly maximizing $\|x\|_D$ is equivalent to maximizing its square. So our goal is to solve

$$\max\left\{\sum_i d_i x_i^2 : x \in P_f\right\} \qquad \text{where } d_i = D_{i,i} > 0.$$

This optimization problem maximizes a convex function over a polytope, and therefore the maximum is attained at one of the vertices. Since any vertex x is a 0-1 vector, it satisfies $x_i^2 = x_i$ for every $i \in [n]$. It follows that

$$\max\left\{\sum_i d_i x_i^2 : x \in P_f\right\} \;=\; \max\left\{\sum_i d_i x_i : x \in P_f\right\}.$$

The latter problem can be solved in polynomial-time (i.e., with a linear number of queries to an independence oracle for the matroid) by the greedy algorithm for finding a maximum weight independent set in a matroid [9].

Theorem 7 *Suppose that f is the rank function of a matroid. Then combining* Approximate-Everywhere *with our exact algorithm for optimizing* $\max\{\,\|x\|_D : x \in P_f\,\}$*, we can construct a $(\sqrt{n}+\epsilon)$-approximation to f in polynomial time.*

We remark that this simple approach of linearizing the objective function (i.e., replacing x_i^2 by x_i) inherently requires the matrix D to be diagonal. For example, the *quadratic spanning tree problem*, defined as $\max\{\,\|x\|_A : x \in P_f\,\}$ where P_f is a graphic matroid polytope and A is an arbitrary symmetric matrix, is NP-hard as it includes the Hamiltonian path problem as a special case [3]. Furthermore, NP-hardness continues to hold under the additional assumption that D is positive definite.

5 General Monotone Submodular Functions

In this section, we present a $1/O(\log n)$-approximate decision procedure for $\max\{\|x\|_D : x \in P_f\}$ for a general monotone submodular function f. Taking squares, we rewrite the problem as:

$$\max\left\{\sum_{i=1}^{n} c_i^2 x_i^2 : x \in P_f\right\}, \tag{9}$$

where we let $c = \mathrm{diag}(D^{1/2})$. Assuming that the ellipsoid $E(D)$ is inscribed in $S(P_f)$, we will either find an $x \in P_f$ for which $\sum_{i=1}^{n} c_i^2 x_i^2 > n+1$ or guarantee that no $x \in P_f$ gives a value greater than $(n+1)/\alpha^2$, where $\alpha = 1/O(\log n)$.

We first consider the case in which all $c_i = 1$, and derive a $(1-1/e)^2$-approximation algorithm for (9). Consider the following greedy algorithm. Let $T_0 = \emptyset$, and for every $k = 1, \cdots, n$, let

$$T_k = \operatorname*{arg\,max}_{S = T_{k-1}\cup\{j\},\, j \notin T_{k-1}} f(S),$$

that is, we repeatedly add the element which gives the largest increase in the submodular function value. Let $\hat{x} \in P_f$ be the vector defined by $\hat{x}(T_k) = f(T_k)$ for $1 \le k \le n$; the fact that $\hat{x}$ is in P_f is a fundamental property of polymatroids. We claim that $\hat{x}$ provides a $(1-1/e)^2$-approximation for (9) when all c_i's are 1.

Lemma 8 *For the solution $\hat{x}$ constructed above, we have*

$$\sum_{i=1}^{n} \hat{x}_i^2 \ge \left(1 - \frac{1}{e}\right)^2 \max\left\{\sum_{i=1}^{n} x_i^2 : x \in P_f\right\}.$$

Proof. Nemhauser, Wolsey and Fisher [36] show that, for every $k \in [n]$, we have

$$f(T_k) \ge \left(1 - \frac{1}{e}\right) \max_{S : |S| = k} f(S).$$

Let $h(k) = f(T_k)$ for $k \in [n]$; because of our greedy choice and submodularity of f, $h(\cdot)$ is concave. Define the monotone submodular function ℓ by $\ell(S) = \frac{e}{e-1}h(|S|)$. The fact that ℓ is submodular comes from the concavity of h. Observe that, for every S, $f(S) \leq \ell(S)$, and therefore, $P_f \subseteq P_\ell$ and

$$\max\left\{\sum_{i=1}^{n} x_i^2 : x \in P_f\right\} \leq \max\left\{\sum_{i=1}^{n} x_i^2 : x \in P_\ell\right\}.$$

By convexity of the objective function, the maximum over P_ℓ is attained at a vertex. But all vertices of P_ℓ are permutations of the coordinates of $\frac{e}{e-1}\hat{x}$ (or are dominated by such vertices), and thus

$$\max\left\{\sum_{i=1}^{n} x_i^2 : x \in P_f\right\} \leq \left(\frac{e}{e-1}\right)^2 \left(\sum_{i=1}^{n} \hat{x}_i^2\right).$$

□

We now deal with the case when the c_i's are arbitrary. First our guarantee that the ellipsoid $E(D)$ is within $S(P_f)$ means that $f(\{i\})e_i$ (where e_i is the ith unit vector) is not in the interior of $E(D)$, i.e. we must have $c_i f(\{i\}) \geq 1$ for all $i \in [n]$. We can also assume that $c_i f(\{i\}) \leq \sqrt{n+1}$. If not, $x = f(\{i\})e_i$ constitutes a vector in P_f with $\sum_j c_j^2 x_j^2 > n+1$. Thus, for all $i \in [n]$, we can assume that $1 \leq c_i f(\{i\}) \leq \sqrt{n+1}$.

To reduce to the case with $c_i = 1$ for all i, consider the linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n : x \to y = (c_1 x_1, \cdots, c_n x_n)$. The problem $\max\{\sum_i c_i^2 x_i^2 : x \in P_f\}$ is equivalent to $\max\{\sum_i y_i^2 : y \in T(P_f)\}$. Unfortunately, $T(P_f)$ is not a polymatroid, but it is contained in the polymatroid P_g defined by:

$$\begin{aligned} g(S) &= \max\left\{\sum_{i \in S} y_i : y \in T(P_f)\right\} \\ &= \max\left\{\sum_{i \in S} c_i x_i : x \in P_f\right\}. \end{aligned}$$

The fact that g is submodular can be derived either from first principles (exploiting the correctness of the greedy algorithm) or as follows. The Lovász extension $\hat{f}$ of f is defined as $f : \mathbb{R}^n \to \mathbb{R} : w \to \max\{w^\mathsf{T} x : x \in P_f\}$ (see Lovász [32] or [12]). It is L-convex, see Murota [34, Prop. 7.25], meaning that, for $w_1, w_2 \in \mathbb{R}^n$, $\hat{f}(w_1) + \hat{f}(w_2) \geq \hat{f}(w_1 \vee w_2) + \hat{f}(w_1 \wedge w_2)$, where $\vee$ (resp. $\wedge$) denotes component-wise max (resp. min). The submodularity of g now follows from the L-convexity of $\hat{f}$ by taking vectors w obtained from c by zeroing out some coordinates.

We can approximately (within a factor $(1-1/e)^2$) compute $\max\{\sum_i y_i^2 : y \in P_g\}$, or equivalently approximate $\max\{\sum_i c_i^2 x_i^2 : x \in T^{-1}(P_g)\}$. The question is how much "bigger" is $T^{-1}(P_g)$ compared to P_f? To answer this question, we perform another polymatroidal approximation, this time of $T^{-1}(P_g)$ and define the submodular function h by:

$$\begin{aligned} h(S) &= \max\left\{\sum_{i \in S} x_i : x \in T^{-1}(P_g)\right\} \\ &= \max\left\{\sum_{i \in S} \frac{1}{c_i} y_i : y \in P_g\right\}. \end{aligned}$$

Again, $h(\cdot)$ is submodular and we can easily obtain a closed form expression for it, see Lemma 11. We have thus sandwiched $T^{-1}(P_g)$ between P_f and P_h: $P_f \subseteq T^{-1}(P_g) \subseteq P_h$. To show that all these polytopes are close to each other, we show the following theorem whose proof is deferred to the full version:

Theorem 9 *Suppose that for all $i \in [n]$, we have $1 \leq c_i f(\{i\}) \leq \sqrt{n+1}$. Then, for all $S \subseteq [n]$, $h(S) \leq \left(2 + \frac{3}{2}\ln(n)\right) f(S)$.*

Our algorithm is now the following. Using the $(1-1/e)^2$-approximation algorithm applied to P_g, we find a vector $\hat{x} \in T^{-1}(P_g)$ such that

$$\sum_i c_i^2 \hat{x}_i^2 \geq \left(1 - \frac{1}{e}\right)^2 \max\left\{\sum_i c_i^2 x_i^2 : x \in T^{-1}(P_g)\right\}.$$

Now, by Theorem 9, we know that $\tilde{x} = \hat{x}/O(\log n)$ is in P_f. Therefore, we have that

$$\begin{aligned} \sum_i c_i^2 \tilde{x}_i^2 &= \frac{1}{O(\log^2(n))} \sum_i c_i^2 \hat{x}_i^2 \\ &\geq \tfrac{1}{O(\log^2(n))} \max\left\{ \textstyle\sum_i c_i^2 x_i^2 : x \in T^{-1}(P_g) \right\} \\ &\geq \tfrac{1}{O(\log^2(n))} \max\left\{ \textstyle\sum_i c_i^2 x_i^2 : x \in P_f \right\}, \end{aligned}$$

giving us the required approximation guarantee.

The lemmas below give a closed form expression for $g(\cdot)$ and $h(\cdot)$; their proofs are used in the proof of Theorem 9. They follow from the fact that the greedy algorithm can be used to maximize a linear function over a polymatroid. Both lemmas apply to any set S after renumbering its indices. For any i and j, we define $[i,j] = \{k \in \mathbb{N} : i \leq k \leq j\}$ and $f(i,j) = f([i,j])$. Observe that $f(i,j) = 0$ for $i > j$.

Lemma 10 *For $S = [k]$ with $c_1 \leq c_2 \leq \cdots \leq c_k$, we have $g(S) = \sum_{i=1}^{k} c_i \left[f(i,k) - f(i+1,k) \right]$.*

Lemma 11 *For $S = [k]$ with $c_1 \leq \cdots \leq c_k$, we have:*

$$\begin{aligned} h(S) = & \sum_{i,j\,:\,1 \leq i \leq j \leq k} \frac{c_i}{c_j} \cdot \Big(f(i,j) - f(i+1,j) \\ & \qquad\qquad - f(i,j-1) + f(i+1,j-1) \Big) \\ = & \sum_{l,m\,:\,1 \leq l \leq m \leq k} (c_l - c_{l-1}) \left(\frac{1}{c_m} - \frac{1}{c_{m+1}} \right) f(l,m). \end{aligned}$$

6 Lower Bound

In this section, we show that approximating a submodular function everywhere requires an approximation ratio of $\Omega\big(\sqrt{n}/\log n\big)$, even when restricting f to be a matroid rank function (and hence monotone). For non-monotone submodular functions, we show that the approximation ratio must be $\Omega\big(\sqrt{n/\log n}\big)$.

The argument has two steps:

- **Step 1.** Construct a family of submodular functions parameterized by natural numbers $\alpha > \beta$ and a set $R \subseteq [n]$ which is unknown to the algorithm.
- **Step 2.** Use discrepancy arguments to determine whether a sequence of queries can determine R. This analysis leads to a choice of α and β.

Step 1. Let $\mathbf{U}$ be the uniform rank-α matroid on $[n]$; its rank function is

$$r_{\mathbf{U}}(S) = \min\left\{ |S|, \alpha \right\}.$$

Now let $R \subseteq [n]$ be arbitrary such that $|R| = \alpha$. We define a matroid $\mathbf{M}_R$ by letting its independent sets be

$$\mathcal{I}_{\mathbf{M}_R} = \left\{ I \subseteq [n] : |I| \leq \alpha \text{ and } |I \cap R| \leq \beta \right\}.$$

This matroid can be viewed as a partition matroid, truncated to rank α. One can check that its rank function is

$$r_{\mathbf{M}_R}(S) = \min\left\{ |S|, \beta + |S \cap \bar{R}|, \alpha \right\}.$$

Now we consider when $r_{\mathbf{U}}(S) \neq r_{\mathbf{M}_R}(S)$. By the equations above, it is clear that this holds iff

$$\beta + |S \cap \bar{R}| < \min\left\{ |S|, \alpha \right\}. \tag{10}$$

Case 1: $|S| \leq \alpha$. Eq. (10) holds iff $\beta + |S \cap \bar{R}| < |S|$, which holds iff $\beta < |S \cap R|$. That inequality together with $|S| \leq \alpha$ implies that $|S \cap \bar{R}| < \alpha - \beta$.

Case 2: $|S| > \alpha$. Eq. (10) holds iff $\beta + |S \cap \bar{R}| < \alpha$. That inequality implies that $|S \cap R| > \beta + (|S| - \alpha) > \beta$.

Our family of monotone functions is

$$\mathcal{F} = \{\, r_{\mathbf{M}_R} : R \subseteq [n], |R| = \alpha \,\} \,\cup\, \{r_{\mathbf{U}}\}\,.$$

Our family of non-monotone functions is

$$\mathcal{F}' = \{\, r_{\mathbf{M}_R} + h : R \subseteq [n], |R| = \alpha \,\} \,\cup\, \{r_{\mathbf{U}} + h\}\,,$$

where h is the function defined by $h(S) = -|S|/2$.

Step 2 (Non-monotone case). Consider any algorithm which is given a function $f \in \mathcal{F}'$, performs a sequence of queries $f(S_1), \ldots, f(S_k)$, and must distinguish whether $f = r_{\mathbf{U}} + h$ or $f = r_{\mathbf{M}_R} + h$ (for some R). For the sake of distinguishing these possibilities, the added function h is clearly irrelevant; it only affects the approximation ratio. By our discussion above, the algorithm can distinguish $r_{\mathbf{M}_R}$ from $r_{\mathbf{U}}$ only if one of the following two cases occurs.

Case 1: $\exists i$ such that $|S_i| \leq \alpha$ and $|S_i \cap R| > \beta$.

Case 2: $\exists i$ such that $|S_i| > \alpha$ and $\beta + |S_i \cap \bar{R}| < \alpha$.

As argued above, if *either* of these cases hold then we have *both* $|S_i \cap R| > \beta$ and $|S_i \cap \bar{R}| < \alpha - \beta$. Thus

$$|S_i \cap R| - |S_i \cap \bar{R}| > 2\beta - \alpha. \tag{11}$$

Now consider the family of sets $\mathcal{A} = \{S_1, \ldots, S_k, [n]\}$. A standard result [1, Theorem 12.1.1] on the discrepancy of $\mathcal{A}$ shows that there exists an R such that

$$\big|\, |S_i \cap R| - |S_i \cap \bar{R}| \,\big| \leq \epsilon \quad \forall i \tag{12a}$$

$$\big|\, |[n] \cap R| - |[n] \cap \bar{R}| \,\big| \leq \epsilon, \tag{12b}$$

where $\epsilon = \sqrt{2n\ln(2k)}$. Eq. (12b) implies that $|R| = n/2 + \epsilon'$, where $|\epsilon'| \leq \epsilon/2$. By definition, $\alpha = |R|$. So if we choose $\beta = n/4 + \epsilon$ then $2\beta - \alpha > \epsilon$. Thus Eq. (11) cannot hold, since it would contradict Eq. (12a). This shows that the algorithm cannot distinguish $f = r_{\mathbf{M}_R} + h$ from $f' = r_{\mathbf{U}} + h$.

The approximation ratio of the algorithm is at most $f'(R)/f(R)$. We have $f'(R) = |R| - |R|/2 = |R|/2$ and $f(R) = \beta - |R|/2 \leq (n/4 + \epsilon) - (n/2 - \epsilon)/2 < 2\epsilon$. This shows that no deterministic algorithm can achieve approximation ratio better than

$$\frac{f'(R)}{f(R)} = \frac{|R|}{4\epsilon} \geq \frac{n/2 - \epsilon}{4\epsilon} = \Omega(\sqrt{n/\log k})$$

Since $k = n^{O(1)}$, this proves the claimed result. If $k = O(n)$ then the lower bound improves to $\Omega(\sqrt{n})$ via a result of Spencer [41].

The construction of the set R in [1, Theorem 12.1.1] is probabilistic: choosing R uniformly at random works with high probability, regardless of the algorithm's queries $S_1, \ldots, S_k$. This implies that the lower bound also applies to randomized algorithms.

Step 2 (Monotone case). In this case, we pick $\alpha \approx \sqrt{n}$ and $\beta = \Omega(\ln k)$. The argument is similar to the non-monotone case except that we cannot apply standard discrepancy results since they do not construct R with $|R| = \alpha \approx \sqrt{n}$. Instead, we derive analogous results using Chernoff bounds. We construct R by picking each element independently with probability $1/\sqrt{n}$. With high probability $|R| = \Theta(\sqrt{n})$. We must now bound the probability that the algorithm succeeds.

Case 1: Given $|S_i| \leq \alpha$, what is $\Pr[\, |S_i \cap R| > \beta \,]$? We have $\mathrm{E}[\, |R \cap S_i| \,] = |S_i|/\sqrt{n} = O(1)$. Chernoff bounds show that $\Pr[\, |R \cap S_i| > \beta \,] \leq \exp(-\beta/2) = 1/k^2$.

Case 2: Given $|S_i| > \alpha$, what is $\Pr\big[\, \beta + |S_i \cap \bar{R}| < \alpha \,\big]$? As observed above, this event is equivalent to $|S_i \cap R| > \beta + (|S_i| - \alpha) =: \xi$. Let $\mu = \mathrm{E}\big[\, |S_i \cap \bar{R}| \,\big] = |S_i|/\sqrt{n}$. Note that

$$\frac{\xi}{\mu} = \frac{\log n}{|S_i|/\sqrt{n}} + \sqrt{n} \cdot \Big(1 - \frac{\alpha}{|S_i|}\Big),$$

which is $\Omega(\log n)$ for any value of $|S_i|$. A Chernoff bound then shows that $\Pr[\, |S_i \cap R| > \xi \,] < \exp(-\xi/2) \leq 1/k^2$.

A union bound shows that none of these events occur with high probability, and thus the algorithm fails to distinguish $r_{\mathbf{M}_R}$ from $r_{\mathbf{U}}$. The approximation ratio of the algorithm is at most $f'(R)/f(R) = \alpha/\beta = \Omega(\sqrt{n}/\log k)$. This lower bound also applies to randomized algorithms, by the same reasoning as in the non-monotone case. Since $k = n^{O(1)}$, this proves the desired result.

7 Applications

7.1 Submodular Load Balancing

Let $f_1, \ldots, f_m$ be monotone submodular functions on the ground set $[n]$. The non-uniform submodular load balancing problem is

$$\min_{V_1,\ldots,V_m} \max_j \; f_j(V_j), \tag{13}$$

where the minimization is over partitions of $[n]$ into $V_1, \ldots, V_m$.

Suppose we construct the approximations $\hat{f}_1, \ldots, \hat{f}_m$ such that

$$\hat{f}_j(S) \le f_j(S) \le g(n) \cdot \hat{f}_j(S) \qquad \forall j \in [m], S \subseteq [n].$$

Furthermore, suppose that each $\hat{f}_j$ is of the form

$$\hat{f}_j(S) = \sqrt{\sum_{i \in S} c_{j,i}},$$

for some non-negative real values $c_{j,i}$. Consider the problem of finding a partition $V_1, \ldots, V_m$ that minimizes $\max_j \hat{f}_j(V_j)$. By squaring, we would like to solve

$$\min_{V_1,\ldots,V_m} \max_j \sum_{i \in V_j} c_{j,i}. \tag{14}$$

This is precisely the problem of scheduling jobs without preemption on non-identical parallel machines, while minimizing the makespan. In deterministic polynomial time, one can compute a 2-approximate solution $X_1, \ldots, X_m$ to this problem [31], which also gives an approximate solution to Eq. (13).

Formally, let $W_1, \ldots, W_m$ be an optimal solution to Eq. (14), let $X_1, \ldots, X_m$ be a solution computed using the algorithm of [31], and let $Y_1, \ldots, Y_m$ be an optimal solution to the original problem in Eq. (13). Then we have $\frac{1}{2} \cdot \max_j \hat{f}_j^2(X_j) \le \max_j \hat{f}_j^2(W_j)$, and thus

$$\frac{1}{\sqrt{2}g(n)} \cdot \max_j f_j(X_j) \le \max_j f_j(Y_j).$$

Thus, the X_j's give a $(\sqrt{2}\, g(n))$-approximate solution to Eq. (13). Applying the algorithm of Section 5 to construct the $\hat{f}_j$'s, we obtain an $O(\sqrt{n} \log n)$-approximation to the non-uniform submodular load balancing problem.

7.2 Submodular Max-Min Fair Allocation

Consider m buyers and a ground set $[n]$ of items. Let $f_1, \ldots, f_m$ be monotone submodular functions on the ground set $[n]$, and let f_j be the valuation function of buyer j. The submodular max-min fair allocation problem is

$$\max_{V_1,\ldots,V_m} \min_j \; f_j(V_j), \tag{15}$$

where the maximization is over partitions of $[n]$ into $V_1, \ldots, V_m$. This problem was studied by Golovin [16] and Khot and Ponnuswami [28]. Those papers respectively give algorithms achieving an $(n-m+1)$-approximation and a $(2m-1)$-approximation. Here we give a $O(n^{\frac{1}{2}} m^{\frac{1}{4}} \log n \log^{\frac{3}{2}} m)$-approximation algorithm for this problem.

The idea of the algorithm is similar to that of the load balancing problem. We construct the approximations $\hat{f}_1, \ldots, \hat{f}_m$

$$\hat{f}_j(S) \le f_j(S) \le g(n) \cdot \hat{f}_j(S) \qquad \forall j \in [m], S \subseteq [n],$$

such that $\hat{f}_j$ is of the form

$$\hat{f}_j(S) = \sqrt{\sum_{i \in S} c_{j,i}},$$

for some non-negative real values $c_{j,i}$. Consider the problem of finding a partition $V_1, \ldots, V_m$ that maximizes $\min_j \hat{f}_j(V_j)$. By squaring, we would like to solve

$$\max_{V_1,\ldots,V_m} \min_j \sum_{i \in V_j} c_{j,i}.$$

This problem is the Santa Claus max-min fair allocation problem, for which Asadpour and Saberi [2] give a $O(\sqrt{m} \log^3 m)$ approximation algorithm. Using this, together with the algorithm of Section 5 to construct the $\hat{f}_j$'s, we obtain an $O(n^{\frac{1}{2}} m^{\frac{1}{4}} \log n \log^{\frac{3}{2}} m)$-approximation for the submodular max-min fair allocation problem.

Acknowledgements

The authors thank Robert Kleinberg for helpful discussions at a preliminary stage of this work, José Soto for discussions on inertial ellipsoids, and Uriel Feige for his help with the analysis of Section 6.

References

[1] N. Alon and J. Spencer. "The Probabilistic Method". Wiley, second edition, 2000.

[2] A. Asadpour and A. Saberi. "An approximation algorithm for max-min fair allocation of indivisible goods". *STOC*, 114–121, 2007.

[3] A. Assad and W. Xu. "The Quadratic Minimum Spanning Tree Problem". *Naval Research Logistics*, **39**, 1992.

[4] K. Ball. "An Elementary Introduction to Modern Convex Geometry". *Flavors of Geometry*, MSRI Publications, 1997.

[5] A. Barvinok. "A Course in Convexity". American Math Society, 2002.

[6] R. G. Bland, D. Goldfarb and M. J. Todd. "The Ellipsoid Method: A Survey". *Operations Research*, **29**, 1981.

[7] S. Boyd and L. Vandenberghe. "Convex Optimization". Cambridge University Press, 2004.

[8] S. Dobzinski and M. Schapira. "An improved approximation algorithm for combinatorial auctions with submodular bidders". *SODA*, 1064–1073, 2006.

[9] J. Edmonds, "Matroids and the Greedy Algorithm", *Mathematical Programming*, **1**, 127–136, 1971.

[10] K. Fan, "On a theorem of Weyl concerning the eigenvalues of linear transformations, II", *Proc. Nat. Acad. Sci.*, 1950.

[11] U. Feige, V. Mirrokni and J. Vondrák, "Maximizing non-monotone submodular functions", *FOCS*, 461–471, 2007.

[12] S. Fujishige, "Submodular Functions and Optimization", volume 58 of *Annals of Discrete Mathematics*. Elsevier, second edition, 2005.

[13] E. Gluskin. "The diameter of the Minkowski compactum is roughly equal to n". Funktsional. Anal. i Prilozhen. 15(1):72–73, 1981.

[14] G. Goel, C. Karande, P. Tripathi, L. Wang. "Approximability of Combinatorial Problems with Multi-agent Submodular Cost Functions", *FOCS*, 2009.

[15] M. Goemans, N. Harvey, R. Kleinberg, V. Mirrokni. "On Learning Submodular Functions". Unpublished Manuscript, 2007.

[16] D. Golovin, "Max-Min Fair Allocation of Indivisible Goods". Technical Report CMU-CS-05-144, 2005.

[17] M. Grötschel, L. Lovász, and A. Schrijver, "Geometric Algorithms and Combinatorial Optimization", Springer Verlag, second edition, 1993.

[18] P. M. Gruber, “Convex and Discrete Geometry”, Springer Verlag, 2009.

[19] O. Güler and F. Gürtina, “The extremal volume ellipsoids of convex bodies, their symmetry properties, and their determination in some special cases”, arXiv:0709.707v1.

[20] S. Iwata, L. Fleischer, and S. Fujishige, “A combinatorial, strongly polynomial-time algorithm for minimizing submodular functions”, *Journal of the ACM*, **48**, 761–777, 2001.

[21] S. Iwata and J. Orlin, “A Simple Combinatorial Algorithm for Submodular Function Minimization”, *SODA*, 2009.

[22] S. Iwata and K. Nagano, “Submodular function minimization under covering constraints”, *FOCS*, 2009.

[23] S. Jegelka, J. Bilmes. “Notes on graph cuts with submodular edge weights”, In Neural Information Processing Society (NIPS) Workshop, Vancouver, Canada, December 2009. Workshop on Discrete Optimization in Machine Learning: Submodularity, Sparsity & Polyhedra (DISCML).

[24] S. Jegelka, J. Bilmes. “Cooperative Cuts: Graph Cuts with Submodular Edge Weights”, Technical Report 189-03-2010, Max Planck Institute for Biological Cybernetics, Tuebingen, 2010.

[25] F. John. “Extremum problems with inequalities as subsidiary conditions”, *Studies and Essays, presented to R. Courant on his 60th Birthday, January 8, 1948*, Interscience, New York, 187–204, 1948.

[26] L. G. Khachiyan. “Rounding of polytopes in the real number model of computation”, *Math of OR*, **21**, 307–320, 1996.

[27] S. Khot, R. Lipton, E. Markakis and A. Mehta. “Inapproximability results for combinatorial auctions with submodular utility functions”, *WINE*, 92–101, 2005.

[28] S. Khot and A. Ponnuswami. “Approximation Algorithms for the Max-Min Allocation Problem”. *APPROX-RANDOM*, 204–217, 2007.

[29] P. Kumar and E. A. Yıldırım, “Minimum-Volume Enclosing Ellipsoids and Core Sets”, *Journal of Optimization Theory and Applications*, **126**, 1–21, 2005.

[30] B. Lehmann, D. J. Lehmann and N. Nisan. “Combinatorial auctions with decreasing marginal utilities”, *Games and Economic Behavior*, **55**, 270–296, 2006.

[31] J. K. Lenstra, D. B. Shmoys and E. Tardos. “Approximation algorithms for scheduling unrelated parallel machines”. *Mathematical Programming*, **46**, 259–271, 1990.

[32] L. Lovász, “Submodular Functions and Convexity”, in A. Bachem et al., eds, Mathematical Programmming: The State of the Art, 235–257, 1983.

[33] J.Matoušek, “Lectures on Discrete Geometry”. Springer, 2002.

[34] K. Murota, “Discrete Convex Analysis”, SIAM Monographs on Discrete Mathematics and Applications, SIAM, 2003.

[35] H. Narayanan, “Submodular Functions and Electrical Networks”, Elsevier, 1997.

[36] G. L. Nemhauser, L. A. Wolsey and M. L. Fisher. “An analysis of approximations for maximizing submodular set functions I”. Mathematical Programming, **14**, 1978.

[37] A. Schrijver, “Combinatorial Optimization: Polyhedra and Efficiency”. Springer-Verlag, 2004.

[38] A. Schrijver, “A combinatorial algorithm minimizing submodular functions in strongly polynomial time”, *Journal of Combinatorial Theory*, Series B, **80**, 346–355, 2000.

[39] R. T. Rockafellar, “Convex Analysis”. Princeton University Press, 1972.

[40] J. Soto, Personal communication, 2008.

[41] J. Spencer, “Six Standard Deviations Suffice”, *Trans. Amer. Math. Soc.*, **289**, 679–706, 1985.

[42] P. Sun and R. M. Freund. “Computation of Minimum Volume Covering Ellipsoids”, *Operations Research*, **52**, 690–706, 2004.

[43] Z. Svitkina and L. Fleischer. “Submodular Approximation: Sampling-Based Algorithms and Lower Bounds”. *FOCS*, 2008.

[44] M. J. Todd. "On Minimum Volume Ellipsoids Containing Part of a Given Ellipsoid". *Math of OR*, 1982.

[45] J. Vondrák. "Optimal Approximation for the Submodular Welfare Problem in the Value Oracle Model". *STOC*, 2008.

[46] L. A. Wolsey. "An Analysis of the Greedy Algorithm for the Submodular Set Covering Problem". *Combinatorica*, **2**, 385–393, 1982.

[47] G. M. Ziegler, "Lectures on Polytopes". Springer-Verlag, 1994.

A Appendix

We give here the proofs that were omitted from the main body.

A.1 Proof of Lemma 2

Lemma 2 *For $A \succ 0$ and $z \in \mathbb{R}^n$ with $l = \|z\|_A^2 \geq n$, let*

$$L(A,z) \;=\; \frac{n}{l}\frac{l-1}{n-1}A + \frac{n}{l^2}\left(1 - \frac{l-1}{n-1}\right)Azz^{\mathsf{T}}A.$$

Then $L(A,z)$ is positive definite, the ellipsoid $E(L(A,z))$ is contained in $\operatorname{conv}(E(A) \cup \{z,-z\})$*, and its volume* $\operatorname{vol} E(L(A,z))$ *equals $\gamma_n(l) \cdot \operatorname{vol} E(A)$ where*

$$\gamma_n(l) \;=\; \sqrt{\left(\frac{l}{n}\right)^n \left(\frac{n-1}{l-1}\right)^{n-1}}.$$

Let us first consider the polar statement to Lemma 2, disregarding the exact definitions of $L(A,z)$ and $\gamma_n(l)$. The polar statement is: if $\|z\|_A > \sqrt{n}$ then there exists an ellipsoid $E(M)$ containing

$$\big(\operatorname{conv}(E(A) \cup \{z,-z\})\big)^*$$

such that $\operatorname{vol} E(M) < \operatorname{vol} E(A^{-1})$. The meaning of this statement is more transparent after the following manipulation.

$$\begin{aligned}
\big(\operatorname{conv}(E(A) \cup \{z,-z\})\big)^* \;&=\; \big(E(A) \cup \{z,-z\}\big)^* && \text{(by (P6))}\\
&=\; E(A)^* \;\cap\; \{z,-z\}^* && \text{(by (P4))}\\
&=\; E(A^{-1}) \;\cap\; \big\{x : -1 \leq z^{\mathsf{T}}x \leq 1\big\}. && (16)
\end{aligned}$$

Thus, the polar of Lemma 2 gives an ellipsoid which contains a section of $E(A)$ determined by centrally symmetric parallel cuts. This is precisely what is used in the ellipsoid method for solving linear programs; see, for example, Grötschel, Lovász and Schrijver [17, p72], Bland, Goldfarb and Todd [6, p1056], and Todd [44]. In fact, Todd derives an exact expression [44, Theorem 2(ii)] for the minimum volume ellipsoid containing (16). After a series of manipulations, one can show that his expression exactly matches our definition of $L(A,z)$, showing that $E(L(A,z))$ is actually the John ellipsoid for $\operatorname{conv}(E(A) \cup \{z,-z\})$.

Proof (of Lemma 2). Let λ and μ be the respective coefficients of A and $Azz^{\mathsf{T}}A$ in $L(A,z)$. For any $n \times n$ matrix H and vectors $u, v \in \mathbb{R}^n$, the Sherman-Morrison formula says that

$$(H + uv^{\mathsf{T}})^{-1} \;=\; H^{-1} - \frac{H^{-1}uv^{\mathsf{T}}H^{-1}}{1 + v^{\mathsf{T}}H^{-1}u},$$

whenever $1 + v^{\mathsf{T}}H^{-1}u \neq 0$. We shall apply this to $L(A,z) = \lambda A + \mu Azz^{\mathsf{T}}A$. Using the facts $l = z^{\mathsf{T}}Az$ and $\lambda + l\mu = n/l$, we obtain

$$L(A,z)^{-1} \;=\; \frac{1}{\lambda}A^{-1} - \frac{\mu l}{\lambda n}zz^{\mathsf{T}}. \tag{17}$$

Now observe that

$$\begin{aligned}
\frac{\mu l}{\lambda n} \;&=\; \frac{(n-1)}{n(l-1)}\left(1 - \frac{l-1}{n-1}\right)\\
&=\; \frac{n-l}{n(l-1)} = \frac{l(n-1)}{n(l-1)} - 1 = \frac{1}{\lambda} - 1.
\end{aligned}$$

Combining this with Eq. (17) shows that

$$L(A,z)^{-1} \;=\; \frac{1}{\lambda}A^{-1} + \left(1 - \frac{1}{\lambda}\right)zz^{\mathsf{T}}. \tag{18}$$

The expression in (18) is clearly positive definite since $A^{-1} \succ 0$, $zz^{\mathsf{T}} \succcurlyeq 0$, and $1 \geq 1/\lambda > 0$. Thus $L(A, z)$ is positive definite.

For an arbitrary $c \in \mathbb{R}^n$, we have

$$c^{\mathsf{T}} L(A, z)^{-1} c = \frac{1}{\lambda} c^{\mathsf{T}} A^{-1} c + \left(1 - \frac{1}{\lambda}\right) (c^{\mathsf{T}} z)^2.$$

Since $\lambda \geq 1$, this implies that

$$\begin{aligned}
&\max \{\, cx : x \in E(L(A, z)) \,\} \\
&\quad = \sqrt{c^{\mathsf{T}} L(A, z)^{-1} c} \\
&\quad \leq \max \left\{ \sqrt{c^{\mathsf{T}} A^{-1} c},\, |c^{\mathsf{T}} z| \right\} \\
&\quad = \max \left\{\, c^{\mathsf{T}} x : x \in \operatorname{conv}(E(A) \cup \{z, -z\}) \,\right\}.
\end{aligned}$$

This shows that $E(L(A, z))^* \supseteq \left(\operatorname{conv}(E(A) \cup \{z, -z\})\right)^*$. By (P8), $E(L(A, z) \subseteq \operatorname{conv}(E(A) \cup \{z, -z\})$.

To compute the volume ratio, we use the following counterpart to the Sherman-Morrison formula.

$$\det(H + uv^{\mathsf{T}}) \;=\; (1 + u^{\mathsf{T}} H^{-1} v) \det H$$

Applying this to $L(A, z)$, we have

$$\det L(A, z) \;=\; \left(1 + \frac{\mu l}{\lambda}\right) \lambda^n \det A \;=\; \left(\frac{n}{l}\right)^n \left(\frac{l-1}{n-1}\right)^{n-1} \det A.$$

Thus, we obtain the required expression for $\gamma_n(l) = \sqrt{\det A / \det L(A, z)}$. $\square$

A.2 Proof of Lemma 3

Lemma 3. *The function $\gamma_n(l)$ given in Lemma 2 satisfies $\gamma_n(n + x) \geq 1 + x^2/(6n^3)$ whenever $0 < x \leq n$.*

Proof (of Lemma 3). Recall that

$$\gamma_n(l) \;=\; \sqrt{\left(\frac{l}{n}\right)^n \left(\frac{n-1}{l-1}\right)^{n-1}}.$$

Then

$$\begin{aligned}
\left(\gamma_n(n + x)\right)^2 \;&=\; \left(\frac{n + x}{n}\right)^n \cdot \left(\frac{n-1}{n + x - 1}\right)^{n-1} \\
&=\; \left(\frac{n + x}{n}\right) \left(\frac{(n + x)(n-1)}{n(n + x - 1)}\right)^{n-1} \\
&=\; \left(\frac{n + x}{n}\right) \left(1 - \frac{x}{n(n + x - 1)}\right)^{n-1} \\
&\geq\; \left(\frac{n + x}{n}\right) \left(1 - \frac{x(n-1)}{n(n + x - 1)}\right) \\
&=\; \frac{(n + x)(n^2 - n + x)}{n^2(n + x - 1)} \\
&=\; 1 + \frac{x^2}{n^2(n + x - 1)} \\
&\geq\; 1 + \frac{x^2}{2n^3}
\end{aligned}$$

The result now follows since $\sqrt{1+y} \geq 1 + y/3$ for $y \in [0,1]$. $\square$

A.3 Proof of Proposition 4

To prove Proposition 4, we will require the following result.

Proposition 12 *Let $A_1, A_2 \succ 0$ and $0 \leq \lambda \leq 1$. Define $B_i = \lambda A_1 + (1-\lambda)A_2$ and $B_o^{-1} = \lambda A_1^{-1} + (1-\lambda)A_2^{-1}$. By convexity of the positive definite cone, both B_i and B_o are positive definite. Then*

$$E(A_1) \cap E(A_2) \;\subseteq\; E(B_i) \;\subseteq\; E(A_1) \cup E(A_2) \tag{19a}$$

$$\log\operatorname{vol} E(B_i) \;\leq\; \lambda \log\operatorname{vol} E(A_1) + (1-\lambda)\log\operatorname{vol} E(A_2). \tag{19b}$$

Also,

$$E(A_1) \cap E(A_2) \;\subseteq\; E(B_o) \;\subseteq\; \operatorname{conv}(E(A_1) \cup E(A_2)) \tag{20a}$$

$$\log\operatorname{vol} E(B_o) \;\geq\; \lambda \log\operatorname{vol} E(A_1) + (1-\lambda)\log\operatorname{vol} E(A_2). \tag{20b}$$

Proof. If $x \in E(A_1) \cap E(A_2)$ then $x^\mathsf{T} A_1 x \leq 1$ and $x^\mathsf{T} A_2 x \leq 1$, implying that $x^\mathsf{T} B_i x \leq 1$ and thus $x \in E(B_i)$. On the other hand, if $x \in E(B_i)$ then $x^\mathsf{T}(\lambda A_1 + (1-\lambda)A_2)x \leq 1$, implying that $\min\left\{x^\mathsf{T} A_1 x,\, x^\mathsf{T} A_2 x\right\} \leq 1$ and thus either $x \in E(A_1)$ or $x \in E(A_2)$. This proves (19a). The fact that $\log\operatorname{vol} E(A) = \log V_n - \frac{1}{2}\log\det A$, together with the log-concavity of the determinant (stated in Lemma 1), implies (19b).

Now we apply (19a) to A_1^{-1} and A_2^{-1}, obtaining

$$\begin{array}{llcccl}
 & E(A_1^{-1}) \cap E(A_2^{-1}) & \subseteq & E(B_o^{-1}) & \subseteq & E(A_1^{-1}) \cup E(A_2^{-1}) \\
\Longrightarrow & (E(A_1)^* \cup E(A_2)^*)^* & \subseteq & E(B_o) & \subseteq & (E(A_1)^* \cap E(A_2)^*)^* \qquad \text{(by (P7) and (P8))} \\
\Longrightarrow & E(A_1) \cap E(A_2) & \subseteq & E(B_o) & \subseteq & \operatorname{conv}(E(A_1) \cup E(A_2)) \qquad \text{(by (P4) and (P9))}
\end{array}$$

We omit the closure operator in the last line because the convex hull of the union of a finite number of compact sets is also compact [39, Corollary 9.8.2]. This proves (20a). The fact that $\log\operatorname{vol} E(A) = \log V_n + \frac{1}{2}\log\det A^{-1}$, together with the log-concavity of the determinant, implies (20b). $\square$

By induction, we obtain the following corollary.

Corollary 13 *Let $A_i \succ 0$ and $\lambda_i \geq 0$ for $1 \leq i \leq k$ with $\sum_i \lambda_i = 1$. Let $L = \operatorname{conv}(\cup_i E(A_i))$. Define*

$$B_i \;=\; \sum_{i=1}^{k} \lambda_i A_i \qquad \textit{and} \qquad B_o^{-1} \;=\; \sum_{i=1}^{k} \lambda_i A_i^{-1}.$$

Then $E(B_i)$ and $E(B_o)$ are ellipsoids both contained within L and they satisfy

$$\log\operatorname{vol} E(B_i) \;\leq\; \sum_{i=1}^{k} \lambda_i \log\operatorname{vol} E(A_i) \;\leq\; \log\operatorname{vol} E(B_o).$$

We are now able to prove Proposition 4.

Proposition 4. *Let K be an axis-aligned convex body, and let $E(A)$ be an ellipsoid inscribed in K. Then the axis aligned ellipsoid $E(B)$ defined by the diagonal matrix $B = (\operatorname{Diag}(A^{-1}))^{-1}$ satisfies $E(B) \subseteq K$ and $\operatorname{vol} E(B) \geq \operatorname{vol} E(A)$.*

Proof. Let $\mathcal{T}$ be the group of all ± 1 diagonal matrices. Then for any matrix M, we have

$$\operatorname{Diag}(M) \;=\; \frac{1}{|\mathcal{T}|}\sum_{T \in \mathcal{T}} T\,M\,T.$$

This holds because, for every i, exactly half the summands negate both the i^{th} row and i^{th} column. Thus

$$B^{-1} \;=\; \operatorname{Diag}(A^{-1}) \;=\; \frac{1}{|\mathcal{T}|}\sum_{T \in \mathcal{T}} T\,A^{-1}\,T \;=\; \frac{1}{|\mathcal{T}|}\sum_{T \in \mathcal{T}} (T\,A\,T)^{-1}.$$

By Corollary 13, $E(B)$ is contained in $\operatorname{conv}\big(\cup_{T\in\mathcal{T}} E(T\,A\,T)\big)$. But $E(T\,A\,T) = T\big(E(A)\big) \subseteq K$, since K is axis aligned. Thus $E(B) \subseteq K$ and

$$\log\operatorname{vol} E(B) \;\geq\; \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \log\operatorname{vol} E(T\,A\,T) \;\geq\; \frac{1}{|\mathcal{T}|}\sum_{T\in\mathcal{T}} \log\operatorname{vol} T\big(E(A)\big) \;=\; \log\operatorname{vol} E(A).$$

This completes the proof. □

A.4 Proof of Lemma 5

Lemma 5. *Let D_0 be the diagonal matrix whose i^{th} diagonal entry is $(n/f(\{i\}))^2$. Then $E(D_0) \subseteq S(P_f)$ and $S(P_f) \subseteq n^2\,E(D_0)$.*

Proof. Any $x \in E(D_0)$ satisfies $\sum_{i=1}^{n} x_i^2 (n/f(\{i\}))^2 \leq 1$ and therefore $|x_i| \leq f(\{i\})/n$ for each $i \in [n]$. This implies that $\sum_{i\in S}|x_i| \leq \max_{i\in S} f(\{i\}) \leq f(S)$. Thus $x \in S(P_f)$, and so the first inclusion is proven.

Any $x \in S(P_f)$ satisfies $|x_i| \leq f(\{i\})$ for every $i \in [n]$. Thus

$$\frac{x^{\mathsf{T}} D_0 x}{n^4} \;=\; \sum_{i=1}^{n} \frac{x_i^2}{n^2 f(\{i\})^2} \;<\; 1.$$

This implies that $x \in n^2\,E(D_0)$, and so the second inclusion is proven. □

A.5 Lemma 10 and Lemma 11

Proof (of Lemma 10). By the greedy algorithm, we have $g(S) = \sum_{i=1}^{k} c_i x_i^*$ where $x_i^* = f(i,k) - f(i+1,k)$ for $i \in [k]$. □

Proof (of Lemma 11). Since $\frac{1}{c_1} \geq \frac{1}{c_2} \geq \cdots \geq \frac{1}{c_k}$, we have

$$h(S) = \sum_{j=1}^{k} \frac{1}{c_j} y_j^*$$

where $y_j^* = g(1,j) - g(1,j-1)$. This means that

$$\begin{aligned}
&h(S)\\
&= \sum_{j=1}^{k} \frac{1}{c_j}\Big(g(1,j) - g(1,j-1)\Big)\\
&= \sum_{j=1}^{k} \frac{1}{c_j}\left(\sum_{i=1}^{j} c_i(f(i,j) - f(i+1,j))\right.\\
&\qquad\qquad \left. - \sum_{i=1}^{j-1} c_i(f(i,j-1) - f(i+1,j-1))\right)\\
&= \sum_{j=1}^{k} \frac{1}{c_j}\left(\sum_{i=1}^{j} c_i(f(i,j) - f(i+1,j))\right.\\
&\qquad\qquad \left. - \sum_{i=1}^{j} c_i(f(i,j-1) - f(i+1,j-1))\right)\\
&= \sum_{i,j:1\leq i\leq j\leq k} \frac{c_i}{c_j}\cdot\Big(f(i,j) - f(i+1,j)
\end{aligned}$$

$$
\begin{aligned}
&\qquad - f(i, j-1) + f(i+1, j-1)\Big) \\
&= \sum_{l,m:1\le l\le m\le k} \left(\frac{c_l}{c_m} - \frac{c_{l-1}}{c_m} - \frac{c_l}{c_{m+1}} + \frac{c_{l-1}}{c_{m+1}}\right) f(l,m) \\
&= \sum_{l,m:1\le l\le m\le k} (c_l - c_{l-1}) \left(\frac{1}{c_m} - \frac{1}{c_{m+1}}\right) f(l,m).
\end{aligned}
$$

□

A.6 Theorem 9

Proof (of Theorem 9). First observe that, for singleton sets $S = \{i\}$, we have $f(S) = h(S)$ (e.g. by definition or from Lemma 11). By scaling f (which scales h as well) and scaling c accordingly (so that $1 \le c_i f(\{i\}) \le \sqrt{n+1}$), we can assume that $\max_{i\in S} f(\{i\}) = 1$.

By submodularity of h, we have $h(S) \le \sum_{i\in S} h(\{i\}) = \sum_{i\in S} f(\{i\}) \le |S| \le |S| f(S)$ by submodularity and monotonicity, and so there is nothing to prove if $|S| \le n \le 2 + \frac{3}{2}\ln n$. In particular, we can assume that $n \ge 5$.

The remainder of the proof is based on the following claim which we will prove shortly.

Claim 14 *If $R = \frac{\max_{i\in S} c_i}{\min_{i\in S} c_i}$ and $f(S) \le |S|/3$ then $h(S) \le (1 + \ln R) f(S)$.*

Assuming the claim, for any set S, let $T = \{i \in S : f(\{i\}) \ge \frac{1}{n-1}\}$. Then

$$
h(S) \le h(T) + \sum_{i\in S\setminus T} h(\{i\}) \le h(T) + 1 \le h(T) + f(S).
$$

If $f(T) > |T|/3$ then we have

$$
\begin{aligned}
h(S) &\le h(T) + f(S) \le 3f(T) + f(S) \\
&\le 4f(S) \le (2 + \frac{3}{2}\ln n) f(S),
\end{aligned}
$$

for $n \ge 5$. On the other hand, if $f(T) \le |T|/3$, we can apply the claim to T to get

$$
h(S) \le (1 + \ln R) f(T) + f(S) \le (2 + \ln R) f(S),
$$

where $R = \frac{\max_{i\in T} c_i}{\min_{i\in T} c_i}$. Since $1 \le c_i f(\{i\}) \le \sqrt{n+1}$ and $\frac{1}{n-1} \le f(\{i\}) \le 1$ for $i \in T$, we get $R \le (n-1)\sqrt{n+1} \le n^{3/2}$, and this gives the right guarantee for the theorem.

We now prove the claim. Fix $S \subseteq [n]$ where $|S| = k$. Assume that $S = [k]$ and that $c_1 \le c_2 \le \cdots c_k$. Furthermore, we know that $c_k = Rc_1$. Scale c so that $c_1 = 1$ (and we do not assume anymore any relationship between c_i and $f(\{i\})$. By submodularity and monotonicity, we know that

$$
f(T) \le \min(|T|, f(S)). \tag{21}
$$

Consider the two equivalent expressions for $h(S)$ given in Lemma 11. Maximize this value over all (not necessarily submodular) functions $f(\cdot)$ satisfying (21) for $T \subset S$ with $f(S)$ fixed at its current value, and over all vectors c with $1 = c_1 \le c_2 \le \cdots \le c_k \le R$. Since the coefficient $(c_l - c_{l-1})(\frac{1}{c_m} - \frac{1}{c_{m+1}})$ of $f(l,m)$ in the second expression is nonnegative, the maximum is attained by the function f^* given by $f^*(T) = \min(|T|, f(S))$ for every $T \subseteq S$. Define $d = \lceil f(S) \rceil$. For this function f^*, we have:

$$
f^*(i,j) = \begin{cases} 0 & \text{if } j - i \le -1 \\ j - i + 1 & \text{if } -1 \le j - i \le d - 2 \\ f(S) & \text{if } d - 1 \le j - i, \end{cases}
$$

and thus, for $1 \leq i \leq j \leq k$, we have:

$$
\begin{aligned}
& f^*(i,j) - f^*(i+1,j) - f^*(i,j-1) + f^*(i+1,j-1) \\
& = \begin{cases} 1 & \text{if } j-i=0 \\ 0 & \text{if } 1 \leq j-i \leq d-2 \\ f(S)-d & \text{if } j-i=d-1 \\ d-1-f(S) & \text{if } j-i=d \\ 0 & \text{if } j-i \geq d+1. \end{cases}
\end{aligned}
$$

The first expression for $h(S)$ given in Lemma 11 now reduces to the following expression for H:

$$H = k - (d - f(S)) \sum_{i=1}^{k-d+1} \frac{c_i}{c_{i+d-1}} - (f(S) - d + 1) \sum_{i=1}^{k-d} \frac{c_i}{c_{i+d}}.$$

We now need to maximize this expression over all vectors c with $1 = c_1 \leq c_2 \leq \cdots \leq c_k \leq R$. By the arithmetic/geometric mean inequality $\frac{1}{n}\sum_{i=1}^{n} a_i \geq (\prod_{i=1}^{n} a_i)^{1/n}$, we get

$$
\begin{aligned}
H \leq k - (d - f(S))(k-d+1) & \left(\prod_{i=1}^{k-d+1} \frac{c_i}{c_{i+d-1}} \right)^{\frac{1}{k-d+1}} \\
-(f(S)-d+1)(k-d) & \left(\prod_{i=1}^{k-d} \frac{c_i}{c_{i+d}} \right)^{\frac{1}{k-d}} \\
= \quad k - (d - f(S))(k-d+1) & \left(\prod_{i=1}^{d-1} \frac{c_i}{c_{i+k-d+1}} \right)^{\frac{1}{k-d+1}} \\
-(f(S)-d+1)(k-d) & \left(\prod_{i=1}^{d} \frac{c_i}{c_{i+k-d}} \right)^{\frac{1}{k-d}}.
\end{aligned}
$$

since this is a telescoping product and $2d \leq k$ whenever $f(S) \leq k/3$. This implies

$$
\begin{aligned}
H \leq\ & k - (d - f(S))(k-d+1)R^{-(d-1)/(k-d+1)} \\
& - (f(S)-d+1)(k-d)R^{-d/(k-d)} \\
\leq\ & k - (d - f(S))(k-d+1)\left(1 - \frac{(\ln R)(d-1)}{k-d+1}\right) \\
& - (f(S)-d+1)(k-d)\left(1 - \frac{(\ln R)d}{k-d}\right) \\
=\ & (1 + \ln R) f(S),
\end{aligned}
$$

where the second inequality follows from $e^{-x} \geq 1 - x$ for any x. This proves the claim and completes the proof of the theorem.

$\square$