Frequency Computation and Bounded Queries

Richard Beigel* William Gasarch† Efim Kinber‡

Yale University Univ. of Maryland Univ. of Delaware

Abstract

There have been several papers over the last ten years that consider the *number of queries* needed to compute a function as a measure of its complexity. The following function has been studied extensively in that light: $\mathrm{F}_a^A(x_1, \ldots, x_a) = A(x_1) \cdots A(x_a)$. We are interested in the complexity (in terms of the number of queries) of *approximating* F_a^A. Let $b \le a$ and let f be any function such that $\mathrm{F}_a^A(x_1, \ldots, x_a)$ and $f(x_1, \ldots, x_a)$ agree on at least b bits. For a general set A we have matching upper and lower bounds that depend on coding theory. These are applied to get exact bounds for the case where A is semirecursive, A is superterse, and (assuming $\mathrm{P} \neq \mathrm{NP}$) $A = \mathrm{SAT}$. We obtain exact bounds when A is the halting problem using different methods.

1 Introduction

The complexity of a function can be measured by the number of queries (to some oracle) needed to compute it. This notion has been studied in both a

*Dept. of Computer Science, Yale University, P.O. Box 208285, New Haven, CT 06520-8285, U.S.A. (email: `beigel@cs.yale.edu`). Supported in part by National Science Foundation grant CCR-8958528.

†Dept. of C.S. and Inst. for Adv. Comp. Stud., Univ. of MD., College Park, MD 20742, U.S.A. Supported in part by NSF grants CCR-8803641 and CCR-9020079 (email: `gasarch@cs.umd.edu`).

‡Univ. of Delaware., Newark, Delaware, 19716, U.S.A. Supported in part by NSF grants CCR-9020079 (email: `kinber@udel.edu`).

recursion-theoretic framework (see for example [5, 11, 17]) and a complexity-theoretic framework (see for example [2, 12, 16]). We give several examples.

1. Let f be the function that, given a graph on n vertices, outputs the number of colors needed to color it. Krentel [16] showed that this function can be computed with $O(\log n)$ queries to SAT in polynomial time but cannot be computed with substantially fewer queries to any oracle in polynomial time (unless P = NP).

2. Let A be a nonrecursive set and $a \in \mathcal{N}$. Let $\#_a^A$ be the function that, given $(x_1, \ldots, x_a)$, returns $|A \cap \{x_1, \ldots, x_a\}|$ (the number of elements that are in A). It is known that there are sets A, X such that $\#_a^A$ can be computed with $\lceil \log(a+1) \rceil - 1$ queries to X. Kummer [17] showed that this is optimal, i.e., if $\#_a^A$ can be computed with $\lceil \log(a+1) \rceil$ queries to some X then A is recursive.

The following functions have been studied extensively in this light:

Definition 1.1 Let $a \in \mathcal{N}$ and $A \subseteq \mathcal{N}$. The function $\mathrm{F}_a^A : \mathcal{N}^a \rightarrow \{0,1\}^a$ is defined as

$$\mathrm{F}_a^A(x_1, \ldots, x_a) = A(x_1) \cdots A(x_a).$$

The function $\#_a^A$ is defined as

$$\#_a^A(x_1, \ldots, x_a) = |A \cap \{x_1, \ldots, x_a\}|.$$

The function F_a^A is interesting because it has a certain intuitive appeal and most lower bounds have reduced to lower bounds for F_a^A. We investigate the complexity of computing an approximation to F_a^A. To do this we define a class of functions $freq_{b,a}^A$ such that every element of $freq_{b,a}^A$ approximates F_a^A.

Notation: If σ, τ are strings of the same length then $\sigma =^a \tau$ means that σ and τ differ in at most a places.

Definition 1.2 Let $a, b \in \mathcal{N}$ be such that $1 \leq b \leq a$, and let $A \subseteq \mathcal{N}$. $freq_{b,a}^A$ is the set of all functions f that map $\mathcal{N}^a$ to $\{0,1\}^a$ such that, for all $x_1, \ldots, x_a$, $f(x_1, \ldots, x_a)$ and $\mathrm{F}_a^A(x_1, \ldots, x_a)$ agree in at least b places (i.e., $f(x_1, \ldots, x_a) =^{a-b} \mathrm{F}_a^A(x_1, \ldots, x_a)$). In prose (though not in theorems) we will informally treat $freq_{b,a}^A$ as just one function: an upper bound on the complexity of $freq^A_{b,a}$ means *at least one* function in $freq^A_{b,a}$ has that complexity (or less), and a lower bound on the complexity of $freq^A_{b,a}$ means that *every* functions in $freq^A_{b,a}$ has that complexity (or greater).

Note: The set $freq^A_{b,a}$ was first defined by Rose [21] and has a long history. For more information see [13].

We investigate the complexity of $freq^A_{b,a}$ for several sets (or types of sets) A and parameters a, b. Our measure of complexity of a function is the number of queries needed to compute it. Most of our results are recursion-theoretic; however, some of our techniques also apply in a polynomial framework.

Information about the complexity of F^A_a will help in our study. However the complexity of $freq^A_{b,a}$ is a harder question. We describe the difference. Assume that, given $(x_1, \ldots, x_a)$, one could produce (the index for) an r.e. set $W \subseteq \{0,1\}^a$ such that $\mathrm{F}^A_a(x_1, \ldots, x_a) \in W$. It has been shown (Lemma 2.4) that the size of W *completely* determines the complexity of F^A_a. Does knowing W help us to compute $freq^A_{b,a}(x_1, \ldots, x_a)$? From W we can obtain W', the set of vectors that differ from elements of W by at most $a - b$ bits. Formally

$$W' = \{\vec{b} : (\exists \vec{c} \in W)[\vec{b} =^{a-b} \vec{c}]\}.$$

It is easy to see that $freq^A_{b,a}(x_1, \ldots, x_a) \in W'$. The complexity of $freq^A_{b,a}$ is completely determined by $|W'|$. Unfortunately it is impossible to determine $|W'|$ from $|W|$. To determine $|W'|$ we need to know the very *structure* of W. This is the key reason that F^A_a is better understood than $freq^A_{b,a}$: the complexity of F^A_a is related to *the cardinality of* W while the complexity of $freq^A_{b,a}$ is related to *the structure of* W. One theme of this paper will be that the more we know about W the better we understand the complexity of $freq^A_{b,a}$.

In Section 3 we prove a general lower bound on the complexity of $freq^A_{b,a}$ (for nonrecursive A). It is based on a general lower bound for $\#^A_a$. In Section 4 we obtain exact bounds for the complexity of $freq^K_{b,a}$. In Section 5 we link the complexity of $freq^A_{b,a}$ to the structure of the set W mentioned above. This will allow us to establish the exact complexity of $freq^A_{b,a}$ for certain sets A. These exact complexities depend on functions from coding theory. In Section 6 we use our proof techniques to obtain results in complexity theory. Assuming $\mathrm{P} \neq \mathrm{NP}$ we determine the exact query complexity of $freq^{\mathrm{SAT}}_{a,b}$.

2 Definitions, Conventions and useful Lemmas

Notation: We use the following notation throughout this paper.

1. $M_0, M_1, \ldots$ is a standard effective list of Turing machines.
2. $M_0^{()}, M_1^{()}, \ldots$ is a standard effective list of oracle Turing machines.
3. W_e is the domain of M_e. Hence $W_0, W_1, \ldots$ is an effective list of all r.e. sets.
4. $K = \{e : M_e(e)\downarrow\}$.
5. If $A \subseteq \mathcal{N}$ then $A' = \{e : M_e^A(e)\downarrow\}$.
6. $D_e = \{i :$ the ith bit of e is $1\}$. Hence $D_0, D_1, \ldots$ is a list of all finite sets.

Convention: Technically M_e takes elements of $\mathcal{N}$ as input and returns elements of $\mathcal{N}$ as output; and $W_e, D_e \subseteq \mathcal{N}$. We will sometimes need to use $\mathcal{N}^a$ (or $\{0,1\}^*$) instead of $\mathcal{N}$. In these cases we implicitly assume that there is a fixed recursive bijection between $\mathcal{N}$ and $\mathcal{N}^a$ ($\{0,1\}^*$) and code elements of $\mathcal{N}^a$ ($\{0,1\}^*$) into $\mathcal{N}$ accordingly.

Definition 2.1 Let $a \in \mathcal{N}$ and let $X \subseteq \mathcal{N}$. $\mathrm{FQ}(a, X)$ is the collection of all total functions g such that g is recursive in X via an algorithm that makes at most a sequential queries to X. $\mathrm{FQC}(a, X)$ is the collection of all functions g such that g is recursive in X via an algorithm $M^{()}$ such that (1) for all x, $M^X(x)$ makes at most a sequential queries to X, and (2) for all x, Y the computation $M^Y(x)$ converges.

The concept of bounded queries is tied to enumerability. Every possible sequence of query answers leads to a possible answer. Hence the fewer queries, the less possible answers.

Definition 2.2 Let $a \in \mathcal{N}$ and f be any total function. f is *a-enumerable* if there exists a recursive function g such that, for all x, $|W_{g(x)}| \leq a$ and $f(x) \in W_{g(x)}$. We denote this by $f \in \mathrm{EN}(a)$. (This concept first appeared in a recursion-theoretic framework in [3]. The name "enumerable" is from [7] where it was defined in a polynomial bounded framework.)

If f is a-enumerable then, given x, we can find $g(x)$ and try to enumerate $W_{g(x)}$ looking for possibilities for $f(x)$. While doing this we do not know when $W_{g(x)}$ will have stopped generating possibilities. The next definition imposes a stronger condition of enumeration. In this scenario we are given an index of a set of possibilities as an index of a finite set. Hence we can obtain all the possibilities and know we have them all.

Definition 2.3 Let $a \in \mathcal{N}$ and f be any total function. f is *strongly a-enumerable* if there exists a recursive function g such that, for all x, $|D_{g(x)}| \leq a$ and $f(x) \in D_{g(x)}$. We denote this by $f \in \mathrm{SEN}(a)$.

Lemma 2.4 ([3, 5]) *Let $a \in \mathcal{N}$ and let f be any function.*

1. $(\exists X)[f \in \mathrm{FQ}(a, X)]$ *iff* $f \in \mathrm{EN}(2^a)$.
2. $(\exists X)[f \in \mathrm{FQC}(a, X)]$ *iff* $f \in \mathrm{SEN}(2^a)$.

In this paper we will prove upper and lower bounds in terms of enumerability (or strong enumerability). Using Lemma 2.4 the reader can obtain corollaries about upper and lower bounds in terms of number of queries.

The following lemma provides a lower bound on the enumerability of $\#_a^A$. We will use it in Theorem 3.1 to obtain a lower bound on $freq_{b,a}^A$.

Lemma 2.5 ([17]) *Let $a \in \mathcal{N}$ and let $A \subseteq \mathcal{N}$. If $\#_a^A \in \mathrm{EN}(a)$ then A is recursive.*

We now exhibit a nonrecursive set A such that if $\frac{b}{a} \leq \frac{1}{2}$ then $freq_{b,a}^A$ is recursive. Since we are interested in how many queries are required to compute $freq_{b,a}^A$ the case where it takes zero queries is not of interest. Hence most of our theorems will assume $\frac{b}{a} > \frac{1}{2}$.

Definition 2.6 [15] A set A is semirecursive if there exists a recursive linear ordering $\sqsubset$ on $\mathcal{N}$ such that A is closed downward under $\sqsubset$.

The following is a folk theorem. It will also be a consequence of Theorem 5.9.

Proposition 2.7 *Assume $\frac{b}{a} \leq \frac{1}{2}$. If A is semirecursive then $freq_{b,a}^A$ is recursive. Hence every tt-degree contains a set A such that $freq_{b,a}^A$ is recursive.*

Proof: Let A be semirecursive via $\sqsubset$. Given $(x_1, \ldots, x_a)$ we may assume $x_1 \sqsubset \cdots \sqsubset x_a$. Since $\mathrm{F}_a^A(x_1, \ldots, x_a) \in \{1^i 0^{a-i} : 0 \leq i \leq a\}$ we have $1^{\lceil \frac{a}{2} \rceil} 0^{\lfloor \frac{a}{2} \rfloor} =^{a-b} \mathrm{F}_a^A(x_1, \ldots, x_a)$. Output $1^{\lceil \frac{a}{2} \rceil} 0^{\lfloor \frac{a}{2} \rfloor}$.

Part 2 follows from part 1 since Jockusch [15] showed that every tt-degree contains a semirecursive set. ▮

It is known that Proposition 2.7 is optimal: if $\frac{b}{a} > \frac{1}{2}$ and $freq_{b,a}^A$ is recursive then A is recursive. This was proven by Trakhtenbrot [22]. We will give an alternative proof (Corollary 3.2).

3 A General Lower Bound for $freq_{b,a}^A$

We prove a general lower bound on the enumerability of $freq_{b,a}^A$ for any nonrecursive A.

Theorem 3.1 *Assume $1 \leq b \leq a$, $\frac{b}{a} > \frac{1}{2}$, and $A \subseteq \mathcal{N}$. If $freq_{b,a}^A \cap \mathrm{EN}(\left\lceil \frac{a+1}{2(a-b)+1} \right\rceil - 1) \neq \emptyset$, then A is recursive.*

Proof: Assume that $f \in freq_{b,a}^A \cap \mathrm{EN}(\left\lceil \frac{a+1}{2(a-b)+1} \right\rceil - 1)$. Let $(x_1, \ldots, x_a) \in \mathcal{N}^a$. Every time a possibility for $f(x_1, \ldots, x_a)$ is generated it yields at most $2(a-b)+1$ possibilities for $\#_a^A(x_1, \ldots, x_a)$. Hence

$$\#_a^A \in \mathrm{EN}(\left(\left\lceil \frac{a+1}{2(a-b)+1} \right\rceil - 1\right)(2(a-b)+1)) \subseteq \mathrm{EN}(a).$$

By Lemma 2.5 A is recursive. ▮

Corollary 3.2 ([22]) *If $\frac{b}{a} > \frac{1}{2}$ and $freq_{b,a}^A$ is recursive, then A is recursive.*

Note: Theorem 3.1 has been obtained independently by Kummer and Stephan [19] using different methods.

4 Exact Bounds for $freq_{b,a}^K$

In this section we determine the *exact* complexity of $freq_{b,a}^K$ in terms of enumerability. In Corollary 5.18 we will determine the *exact* complexity of $freq_{b,a}^K$ in terms of strong enumerability. It is known that $\#_a^K(x_1,\ldots,x_a)$ completely determines F_a^K. Hence the structure of the set of possibilities for F_a^K is well understood. This is why we are able to obtain exact bounds.

Theorem 4.1 *If* $1 \le b \le a$ *then* $freq_{b,a}^K \cap \mathrm{EN}(\lceil \frac{a+1}{(a-b)+1} \rceil) \neq \emptyset$.

Proof: Given $(x_1,\ldots,x_a)$ we show how to enumerate $\le \lceil \frac{a+1}{(a-b)+1} \rceil$ possibilities such that one of them is $=^{a-b} \mathrm{F}_a^K(x_1,\ldots,x_a)$.

Let $k = \lceil \frac{a+1}{(a-b)+1} \rceil$. Since $b \ge 1$ we have $k \ge 1$. Let $I_1,\ldots,I_k$ be intervals of length $\le a-b+1$ that partition $\{0,\ldots,a\}$.

For each interval $I=[c,d]$ we enumerate a possibility that is based on the belief that $\#_a^K(x_1,\ldots,x_a) \in [c,d]$. By dovetailing these computations we enumerate $\le k$ possibilities.

For interval $I=[c,d]$ we do the following. If $c=0$ then output $(0,\ldots,0)$. If $c>0$ then simultaneously run all of $M_{x_1}(x_1),\ldots,M_{x_a}(x_a)$ until exactly c of them halt (this need not happen). Output a string that indicates that these c programs are in K and no other programs are in K.

We show that if $\#_a^K(x_1,\ldots,x_a) \in I = [c,d]$ then the possibility associated to I is correct. Clearly the c 1's are correct. Since there are at most d programs in K, at least $a-d$ of the 0's are correct. Hence at least $c+a-d = a+(c-d) = a+1-|I| \ge a+1-(a-b+1) = b$ bits are correct. ∎

Note: By Lemma 2.4, $(\exists X)[freq_{b,a}^K \cap \mathrm{FQ}(\lceil \log \frac{a+1}{(a-b)+1} \rceil, X) \neq \emptyset]$. The oracles is unspecified. In this case we can do just as well with oracle K: by a truncated binary search, $freq_{b,a}^K \cap \mathrm{FQ}(\lceil \log \frac{a+1}{(a-b)+1} \rceil, K) \neq \emptyset$.

The enumeration procedure used in Theorem 4.1 is not a strong enumeration. In Section 5 we show that a strong enumeration for $freq_{b,a}^K$ requires many more possibilities than an enumeration.

We show that the above bound is tight.

Theorem 4.2 *If* $1 \le b \le a$ *then* $freq_{b,a}^K \cap \mathrm{EN}(\lceil \frac{a+1}{(a-b)+1} \rceil - 1) = \emptyset$.

Proof: Assume $f \in freq^K_{b,a} \cap \mathrm{EN}(\lceil \frac{a+1}{(a-b)+1} \rceil - 1) \neq \emptyset$. Assume $f \in \mathrm{EN}(\lceil \frac{a+1}{(a-b)+1} \rceil - 1)$ via g. We create a programs x_i that conspire to cause

$$(\forall \vec{b} \in W_{g(x_1,\ldots,x_a)})[\neg(\vec{b} =^{a-b} \mathrm{F}^K_a(x_1,\ldots,x_a))].$$

We plan to have different blocks of programs invalidate different elements of $W_{g(x_1,\ldots,x_a)}$. Let $k = \lceil \frac{a+1}{(a-b)+1} \rceil - 1$. Since $b \geq 1$ we have $k \geq 1$. Let $J_1, \ldots, J_k$ be intervals of length $\geq a-b+1$ that partition $\{0,\ldots,a\}$.

By the a-ary recursion theorem we can assume that x_i has access to the numbers $\{x_1,\ldots,x_a\}$.
ALGORITHM FOR x_i

1. Let j be such that $i \in J_j$ (if no such j exists then diverge).
2. Enumerate $W_{g(x_1,\ldots,x_a)}$ until j elements appear (this step might not terminate). Let that jth element be $\vec{b} = b_1 \cdots b_a$.
3. If $b_i = 0$ then converge. If $b_i = 1$ then diverge.

END OF ALGORITHM

For all j, $1 \leq j \leq k$, if $W_{g(x_1,\ldots,x_a)}$ has the jth element $\vec{b}$, then $\vec{b}$ and $\mathrm{F}^K_a(x_1,\ldots,x_a)$ differ on the bits specified by J_j. Hence they differ on at least $a-b+1$ places, so $(\forall \vec{b} \in W_{g(x_1,\ldots,x_a)})[\neg(\vec{b} =^{a-b} \mathrm{F}^K_a(x_1,\ldots,x_a))]$. ∎

5 Exact Bounds for $freq^A_{b,a}$

In this section we prove a general theorem relating the complexity of $freq^A_{b,a}$ to the structure of the set of possible values for F^A_a. We then apply this theorem to semirecursive sets, joins of semirecursive sets, and superterse sets.

We will need some definitions from coding theory.

Definition 5.1 Let $a, r \in \mathcal{N}$. Let $z \in \{0,1\}^a$. The *closed ball of radius* r *centered at* z is the set $B(z,r) = \{y \in \{0,1\}^a : y =^r z\}$. If $D \subseteq \{0,1\}^a$ then D *is covered by* k *balls of radius* r means that there exist $z_1,\ldots,z_k$ such that $D \subseteq \bigcup_{i=1}^k B(z_i,r)$.

Definition 5.2 Let $a, r \in \mathcal{N}$ and $D \subseteq \{0,1\}^a$. $k(D,r)$ is the minimal number j such that D can be covered by j balls of radius r. The quantity $k(\{0,1\}^a, r)$ is denoted by $k(a,r)$.

Let D be a set of possibilities for $\mathrm{F}_a^A(x_1,\ldots,x_a)$ such that $k(D, a-b) = j$. Let $\vec{b} \in D$ be the correct possibility. Let $\vec{c}$ be the center of the ball that contains $\vec{b}$. Since $\vec{b}$ and $\vec{c}$ differ on at most $a-b$ places they must agree on at least b places. Hence $\vec{c}$ is a suitable value of $freq_{b,a}^A$.

The quantity $k(a,r)$ is known as *the covering number*. It has been studied extensively (see [8, 9, 10, 14, 23]). No exact formula is known for it, however we present some known estimates.

Fact 5.3 *Let* $S_{a,r} = \sum_{i=0}^{r} \binom{a}{i}$.

1. $\frac{2^a}{S_{a,r}} \leq k(a,r) \leq \frac{2^a}{S_{a,r}}(1+\log S_{a,r})$ *([8, Theorem 3]. (Better lower bounds are known [23, Theorem 10].)*
2. $k(r+1,r) = k(r+2,r) = \cdots = k(2r+2,r) = 2$ *([10, Theorem 14]).*
3. $k(2r+3,r) = 3$, *and* $7 \leq k(2r+4,r) \leq 12$ *([10, Theorem 14]).*

Definition 5.4 Let $a, r \in \mathcal{N}$ and $\mathcal{D} \subseteq 2^{\{0,1\}^a}$. We define $k(\mathcal{D}, r)$ to be $\max\{k(D,r) : D \in \mathcal{D}\}$.

We need to define the notions of $\mathcal{D}$-verbose and strongly $\mathcal{D}$-verbose in order to state our main result. Note that every set is strongly $2^{\{0,1\}^a}$-verbose.

Definition 5.5 Let $a \in \mathcal{N}$. Let $\mathcal{D} \subseteq 2^{\{0,1\}^a}$. A set A is *$\mathcal{D}$-verbose* if there is a recursive function g such that, for all $x_1,\ldots,x_a$, $W_{g(x_1,\ldots,x_a)} \in \mathcal{D}$ and $\mathrm{F}_a^A(x_1,\ldots,x_a) \in W_{g(x_1,\ldots,x_a)}$. A set A is *strongly $\mathcal{D}$-verbose* if there is a recursive function g such that, for all $x_1,\ldots,x_a$, $D_{g(x_1,\ldots,x_a)} \in \mathcal{D}$ and $\mathrm{F}_a^A(x_1,\ldots,x_a) \in D_{g(x_1,\ldots,x_a)}$.

The following theorem provides for any $A \subseteq \mathcal{N}$: (1) matching upper and lower bounds for the strong enumerability of $freq_{b,a}^A$, and (2) lower bounds for the enumerability of $freq_{b,a}^A$. All results in this paper, except those involving $freq_{b,a}^K$, will follow from it.

Theorem 5.6 *Assume $1 \le b \le a$ and $A \subseteq \mathcal{N}$. All $\mathcal{D}$ mentioned in this theorem are understood to be subsets of $2^{\{0,1\}^a}$. For all k the following hold.*

1. *If there exists $\mathcal{D}$ such that A is strongly $\mathcal{D}$-verbose and $k(\mathcal{D}, a-b) \le k$ then $freq^A_{b,a} \cap \mathrm{SEN}(k) \neq \emptyset$.*
2. *If $freq^A_{b,a} \cap \mathrm{SEN}(k) \neq \emptyset$ then there exists $\mathcal{D}$ such that A is strongly $\mathcal{D}$-verbose and $k \ge k(\mathcal{D}, a-b)$.*
3. *If $freq^A_{b,a} \cap \mathrm{EN}(k) \neq \emptyset$ then there exists $\mathcal{D}$ such that A is $\mathcal{D}$-verbose and $k \ge k(\mathcal{D}, a-b)$.*

Proof:
(1) Assume A is strongly $\mathcal{D}$-verbose via g. Given $(x_1, \ldots, x_a)$ we strongly enumerate $\le k$ possibilities one of which must be $=^{a-b} \mathrm{F}^A_a(x_1, \ldots, x_a)$. Find $D = D_{g(x_1,\ldots,x_a)}$. Find a set of vectors $\{\vec{b}_1, \ldots, \vec{b}_k\}$ such that $D \subseteq \bigcup_{i=1}^k B(\vec{b}_i, a-b)$. (Such vectors exist since $k(\mathcal{D}, a-b) \le k$.) Enumerate $\vec{b}_1, \ldots, \vec{b}_k$ as possibilities. Since $\mathrm{F}^A_a(x_1, \ldots, x_a) \in D$

$$(\exists i)[\mathrm{F}^A_a(x_1, \ldots, x_a) \in B(\vec{b}_i, a-b)]$$

so

$$(\exists i)[\mathrm{F}^A_a(x_1, \ldots, x_a) =^{a-b} \vec{b}_i].$$

(2) Assume $freq^A_{b,a} \cap \mathrm{SEN}(k) \neq \emptyset$. Then there exist k total recursive functions $p_1, \ldots, p_k$ such that $(\forall x_1, \ldots, x_a)(\exists i)[p_i(x_1, \ldots, x_a) =^{a-b} \mathrm{F}^A_a(x_1, \ldots, x_a)]$. Let

$$\begin{aligned} D_{g(x_1,\ldots,x_a)} &= \bigcup_{i=1}^k B(p_i(x_1, \ldots, x_a), a-b) \\ \mathcal{D} &= \{D_{g(x_1,\ldots,x_a)} : x_1, \ldots, x_a \in \mathcal{N}\} \end{aligned}$$

Clearly A is strongly $\mathcal{D}$-verbose. Since every element of $\mathcal{D}$ is a union of k balls of radius $a-b$, $k \ge \max\{k(D, a-b) : D \in \mathcal{D}\}$.
(3) Similar to the proof of part 2. ∎

Theorem 5.6 yields matching upper and lower bounds; however they are not readily computable. The following lemma will be helpful in computing them.

Lemma 5.7 *Let $a, r \in \mathcal{N}$ and $A \subseteq \mathcal{N}$.*

1. *If there exists $\mathcal{D}$ such that A is strongly $\mathcal{D}$-verbose and $k = k(\mathcal{D}, r)$ then* $\#_a^A \in \mathrm{SEN}(k(2r+1))$.

2. *If there exists $\mathcal{D}$ such that A is (strongly) $\mathcal{D}$-verbose then* F_a^A *is (strongly)* $\max\{|D| : D \in \mathcal{D}\}$*-enumerable.*

Proof:

(1) Assume A is strongly $\mathcal{D}$-verbose via g. We show how to $k(2r+1)$-enumerate $\#_a^A$. On input $(x_1, \ldots, x_a)$ find $D = D_{g(x_1,\ldots,x_a)}$. We know D can be covered by k balls of radius r. Let $\vec{b}_1, \ldots, \vec{b}_k$ be the centers of those balls. Let a_i be the number of 1's in $\vec{b}_i$. Enumerate

$$\{a_i + a : 1 \leq i \leq k \text{ and } -r \leq a \leq r\}.$$

These are the $k(2r+1)$ numbers one of which must be $\#_a^A(x_1, \ldots, x_a)$.

(2) This follows from the definition of (strongly) $\mathcal{D}$-verbose.

∎

Note: Kummer and Stephan [18, Corollary 4.3,4.4] have found a different connection between covering numbers and $freq_{b,a}^A$. Let $\Omega(b,a) = \{A : freq_{b,a}^A$ is recursive $\}$. They have shown the following.

1. $(\forall a \geq 2)(\exists A, A$ 2-r.e. $)[A \in \Omega(1, \lceil \log(k(a,1)+1) \rceil) - \Omega(2,a)]$.

2. $(\forall b \geq 2)(\exists A, A$ r.e. $)[A \in \Omega(1, 2^b - b) - \Omega(2, 2^b - 1)]$.

5.1 Semirecursive Sets

We established matching upper and lower bounds for $freq_{b,a}^A$ when A is semirecursive using Proposition 2.7 and Theorem 3.1. Here we give an alternative proof using our general theorem.

Lemma 5.8 *Let $D = \{1^i 0^{a-i} : 0 \leq i \leq a\}$, and let $0 \leq r \leq a$. Then* $k(D,r) = \left\lceil \frac{a+1}{2r+1} \right\rceil$.

Proof: Let $k = \left\lceil \frac{a+1}{2r+1} \right\rceil$. For $1 \le i \le k-1$ let $z_i = 1^{(2i-1)r+i-1}0^{a-(2i-1)r-i+1}$, and let $z_k = 1^{a-r}0^r$. It is easy to check that $D \subseteq \bigcup_{i=1}^{k} B(z_i, r)$. Hence $k(D, r) \le k$.

If $\le k-1$ balls of radius r are used then $\le (k-1)(2r+1) \le a$ elements are covered. Hence $k(D, r) \ge k$.

Combining the inequalities we obtain $k(D, r) = k$. ∎

Theorem 5.9 *Assume $1 \le b \le a$, A is a nonrecursive semirecursive set, and $k = \left\lceil \frac{a+1}{2(a-b)+1} \right\rceil$. Then $freq^A_{b,a} \cap \mathrm{SEN}(k) \ne \emptyset$ but $freq^A_{b,a} \cap \mathrm{SEN}(k-1) = \emptyset$. Note that if $\frac{b}{a} \le \frac{1}{2}$ then $k = 1$ so $freq^A_{b,a} \cap \mathrm{EN}(1) \ne \emptyset$, hence some function in $freq^A_{b,a}$ is recursive.*

Proof:

Let A be a semirecursive set with ordering $\sqsubset$. Let $D = \{1^i 0^{a-i} : 0 \le i \le a\}$. Let $\mathcal{D}$ be the singleton set $\{D\}$. Semirecursive sets are strongly $\mathcal{D}$-verbose: on input $(x_1, \ldots, x_a)$ (assume $x_1 \sqsubset \cdots \sqsubset x_a$) the only possibilities for $\mathrm{F}^A_a(x_1, \ldots, x_a)$ are $1^i 0^{a-i}$ where $0 \le i \le a$.

By Theorem 5.6 $freq^A_{b,a} \cap \mathrm{SEN}(k(D, a-b)) \ne \emptyset$. Since $0 \le a-b \le a$ we can apply Lemma 5.8 with $r = a-b$. Hence $freq^A_{b,a} \cap \mathrm{SEN}(k) \ne \emptyset$.

Assume, by way of contradiction, that $freq^A_{b,a} \cap \mathrm{SEN}(k-1) \ne \emptyset$. By Theorem 5.6 there exists $\mathcal{D}$ such that A is strongly $\mathcal{D}$-verbose and $k(\mathcal{D}, a-b) = k-1$. By Lemma 5.7 $\#^A_a \in \mathrm{EN}((k-1)(2(a-b)+1)) \subseteq \mathrm{EN}(a)$. By Lemma 2.5 A is recursive. ∎

5.2 Joins of Semirecursive Sets

In this section we obtain an upper bound on the complexity of $freq^A_{b,a}$ when A is the join of several semirecursive sets. No lower bound is known in the general case; however there are particular sets A of this type for which the lower bound is tight.

Definition 5.10 If D_1 and D_2 are sets of strings then

$$D_1 \cdot D_2 = \{\sigma\tau : \sigma \in D_1 \text{ and } \tau \in D_2\}.$$

Definition 5.11 If $A_1, A_2 \subseteq \mathcal{N}$ then

$$A_1 \oplus A_2 = \{2x : x \in A_1\} \cup \{2x + 1 \mid x \in A_2\}.$$

Lemma 5.12 *Let $a_1, \ldots, a_q, r_1, \ldots, r_q$ and $D_1, \ldots, D_q$ be such that $D_i \subseteq \{0,1\}^{a_i}$ for all i. Then*

$$k(D_1 \cdot D_2 \cdots D_q, r) \leq \min\left\{\prod_{i=1}^{q} k(D_i, r_i) : (\forall i)[r_i \geq 1] \text{ and } \sum_{i=1}^{q} r_i = r\right\}.$$

Proof: We prove this for $q = 2$. The general case is similar. Let $r = r_1 + r_2$ be some partition of r into nonzero parts. Let k_1 and k_2 be such that $k(D_i, r_i) = k_i$. Let $y_1, \ldots, y_{k_1}, z_1, \ldots, z_{k_2}$ be such that $D_1 \subseteq \bigcup_{i=1}^{k_1} B(y_i, r_1)$ and $D_2 \subseteq \bigcup_{i=1}^{k_2} B(z_i, r_2)$. It is easy to see that

$$D_1 \cdot D_2 \subseteq \bigcup_{i=1}^{k_1} \bigcup_{j=1}^{k_2} B(y_i \cdot z_j, r_1 + r_2).$$

Hence $k(D_1 \cdot D_2, r) \leq k_1 k_2 = k(D_1, r_1)k(D_2, r_2)$. Since this holds for any nonzero partition $r = r_1 + r_2$ we can take r_1, r_2 that results in the minimal $k(D_1, r_1)k(D_2, r_2)$. ∎

Theorem 5.13 *Assume $1 \leq b \leq a$, $\frac{b}{a} > \frac{1}{2}$, and $q \geq 1$. Let $A_1, \ldots, A_q$ be semirecursive sets. Let $A = A_1 \oplus \cdots \oplus A_q$.*

1. *$freq_{b,a}^{A} \cap \mathrm{SEN}(k) \neq \emptyset$ where k is defined as follows.*

$$k = \max\{\min\left\{\prod_{i=1}^{q} \left\lceil \frac{a_i + 1}{r_i + 1} \right\rceil : \sum_{i=1}^{q} r_i = a - b\right\} : \sum_{j=1}^{q} a_i = a\}.$$

2. *If q divides both a and b then $freq_{b,a}^{A} \cap \mathrm{EN}(\left(\left\lceil \frac{a+q}{a-b+q} \right\rceil\right)^q) \neq \emptyset$.*

Proof:
1) For any a', $0 \le a' \le a$, let $E^{a'} = \{1^i 0^{a'-i} : 0 \le i \le a'\}$. Note that A is strongly $\mathcal{D}$-verbose where $\mathcal{D} = \{\prod_{i=1}^q E^{a_i} : \sum_{j=1}^q a_i = a\}$. By Theorem 5.6 $freq_{b,a}^A \cap \mathrm{SEN}(k) \neq \emptyset$ where

$$k = \max\{k\left(\prod_{i=1}^q E^{a_i}, a-b\right) : \sum_{j=1}^q a_i = a\}.$$

By Lemmas 5.12 and 5.8

$$\begin{aligned} k(\textstyle\prod_{i=1}^q E^{a_i}, a-b) &\le \min\{\textstyle\prod_{i=1}^q k(E^{a_i}, r_i) : \sum_{i=1}^q r_i = a-b\} \\ &\le \min\{\textstyle\prod_{i=1}^q \left\lceil \frac{a_i+1}{r_i+1} \right\rceil : \sum_{i=1}^q r_i = a-b\} \end{aligned}$$

Putting this all together we obtain that $freq_{b,a}^A \cap \mathrm{SEN}(k) \neq \emptyset$. where

$$k = \max\{\min\left\{\prod_{i=1}^q \left\lceil \frac{a_i+1}{r_i+1} \right\rceil : \sum_{i=1}^q r_i = a-b\right\} : \sum_{j=1}^q a_i = a\}.$$

(2) If q divides b and a, then q divides $a-b$. In this case the internal min occurs when all r_i's are $\frac{a-b}{q}$. Hence

$$k = \max\left\{\prod_{i=1}^q \left\lceil \frac{a_i+1}{\frac{a-b}{q}+1} \right\rceil : \sum_{j=1}^q a_i = a\right\}.$$

The max occurs when all a_i's are $\frac{a}{q}$. When this occurs

$$k = \prod_{i=1}^q \left\lceil \frac{\frac{a}{q}+1}{\frac{a-b}{q}+1} \right\rceil = \left(\left\lceil \frac{\frac{a}{q}+1}{\frac{a-b}{q}+1} \right\rceil\right)^q = \left(\left\lceil \frac{a+q}{a-b+q} \right\rceil\right)^q.$$

∎

There are semirecursive sets $A_1, \ldots, A_b$ where the upper bound from Theorem 5.13 is an overestimate; for example, if $A_1 = \cdots = A_b$ then $\mathrm{F}_a^{A_1 \oplus \cdots \oplus A_b} \in \mathrm{SEN}(a+1)$. However, Theorem 5.13 is optimal for the general case:

Theorem 5.14 *Let a, b, q, k be as in Theorem 5.13. There exist sets $A_1, \ldots, A_q$ and $A = A_1 \oplus \cdots \oplus A_q$ such that $freq_{b,a}^A \cap \mathrm{EN}(k-1) = \emptyset$.*

Proof:

This can be proven by a straightforward diagonalization similar to [11, Appendix]. ∎

5.3 Superterse and Weakly Superterse Sets

Definition 5.15 [5] A set A is *superterse* if $(\forall n)(\forall X)[\mathrm{F}_n^A \notin \mathrm{FQ}(n-1, X)]$. A set A is *weakly superterse* if $(\forall n)(\forall X)[\mathrm{F}_n^A \notin \mathrm{FQC}(n-1, X)]$.

Lemma 5.16 ([4]) *Let $A \subseteq \mathcal{N}$.*

1. *If there exists a such that $\mathrm{F}_a^A \in \mathrm{EN}(2^a - 1)$ ($\mathrm{F}_a^A \in \mathrm{SEN}(2^a - 1)$), then there exists a constant c such that $(\stackrel{\infty}{\forall} n)[\mathrm{F}_n^A \in \mathrm{EN}(n^c)]$ ($\mathrm{F}_n^A \in \mathrm{SEN}(n^c)$).*
2. *Assume A is (weakly) superterse. For all n, $\mathrm{F}_n^A \notin \mathrm{EN}(2^n - 1)$. ($\mathrm{F}_n^A \notin \mathrm{SEN}(2^n - 1)$). This follows from part 1 and Lemma 2.4.*

(Note that [4] shows a complexity-theoretic version of this lemma, but the proof can be easily modified to obtain this lemma.)

Theorem 5.17 *Assume $1 \le b \le a$, $\frac{b}{a} > \frac{1}{2}$, and $A \subseteq \mathcal{N}$.*

1. *$freq_{b,a}^A \cap \mathrm{SEN}(k(a, a-b)) \neq \emptyset$.*
2. *If A is superterse then $freq_{b,a}^A \cap \mathrm{EN}(k(a, a-b) - 1) = \emptyset$.*
3. *If A is weakly superterse then $freq_{b,a}^A \cap \mathrm{SEN}(k(a, a-b) - 1) = \emptyset$.*

Proof:

(1) This follows from Theorem 5.6.

(2) Let A be superterse. Assume, by way of contradiction, that $freq_{b,a}^A \cap \mathrm{EN}(k(a, a-b) - 1) \neq \emptyset$. By Theorem 5.6 there exists $\mathcal{D}$ such that A is $\mathcal{D}$-verbose and $k(\mathcal{D}, a-b) = k(a, a-b) - 1$. Hence, for every $D \in \mathcal{D}$, $k(D, a-b) \le k(a, a-b) - 1$ so $|D| \le 2^a - 1$. By Lemma 5.7, $\mathrm{F}_a^A \in \mathrm{EN}(2^a - 1)$. By Lemma 5.16, A is not superterse.

(3) Similar to part 2.

∎

Corollary 5.18 *Assume* $1 \le b \le a$.

1. $freq^K_{b,a} \cap \mathrm{SEN}(k(a, a-b)) \neq \emptyset$ *but* $freq^K_{b,a} \cap \mathrm{SEN}(k(a, a-b) - 1) = \emptyset$.

2. *For every nonrecursive set* A, $freq^{A'}_{b,a} \cap \mathrm{EN}(k(a, a-b)) \neq \emptyset$ *but* $freq^{A'}_{b,a} \cap \mathrm{EN}(k(a, a-b) - 1) = \emptyset$. *(Recall that* A' *is the halting problem relative to* A.*)*

3. *Every nonzero truth-table degree contains a set* A *such that* $freq^A_{b,a} \cap \mathrm{SEN}(k(a, a-b)) \neq \emptyset$ *but* $freq^A_{b,a} \cap \mathrm{EN}(k(a, a-b) - 1) = \emptyset$.

Proof: By [11, Theorem 23], K is weakly superterse. By [5, Theorem 16], for all nonrecursive A, A' is superterse. By [5, Theorem 14], every nonzero tt-degree contains a superterse set. ∎

Theorems 4.1 and Corollary 5.18 offer an interesting contrast. We obtain the exact complexity of $freq^K_{b,a}$ via (1) algorithms that need not halt if a different oracle is used, (2) algorithms that halt regardless of the oracle. The following table shows that the difference in complexity is small when $b \le \frac{a}{2} + 2$, but is exponentially large when $a - b$ is constant. We show how the table is derived and impose bounds as to when the rows of the table apply. The rule $b \le a$ always applies.

1. If $2b = a+4$ then $a = 2(a-b)+4$, hence $k(a, a-b) = k(2(a-b)+4, a-b)$. If $a - b \ge 1$ then by Fact 5.3 $7 \le k(2(a-b)+4, a-b) \le 12$. If $a - b \ge 1$ then by Corollary 5.18 and Lemma 2.4 the optimal number of queries needed to compute $freq^K_{b,a}$ is either 3 or 4. This derivation only applies to $a - b \ge 1$, hence the first row of the table may be excluded in the case $a = 4$. In that case we are considering $freq^K_{4,4}$ which is the same as F^K_4. By [11, Theorem 23] $\mathrm{F}^K_4 \in \mathrm{FQC}(4, K) - \mathrm{FQC}(3, K)$. Hence the information in the table is still valid.

2. If $2b = a+3$ then $a = 2(a-b)+3$, hence $k(a, a-b) = k(2(a-b)+3, a-b)$. If $a-b \ge 1$ then by Fact 5.3 $k(2(a-b)+3, a-b) = 3$. If $a-b \ge 1$ then by Corollary 5.18 and Lemma 2.4 the optimal number of queries needed to compute $freq^K_{b,a}$ is 2. This derivation only applies to $a - b \ge 1$, hence the second row of the table may be excluded in the case $a = 3$. In that case we are considering $freq^K_{3,3}$ which is the same as F^K_3. By [11, Theorem 23] $\mathrm{F}^K_3 \in \mathrm{FQC}(3, K) - \mathrm{FQC}(3, K)$. Hence the information in the table is not valid for $a = 3$.

3. If $2b = a+2$ then $a = 2(a-b)+2$, hence $k(a, a-b) = k(2(a-b)+2, a-b)$. If $a-b \geq 1$ then by Fact 5.3 $k(2(a-b)+2, a-b) = 2$. If $a-b \geq 1$ then by Corollary 5.18 and Lemma 2.4 the optimal number of queries needed to compute $freq^K_{b,a}$ is 1. This derivation only applies to $a - b \geq 1$, hence the third row of the table may be excluded in the case $a = 2$. In that case we are considering $freq^K_{2,2}$ which is the same as F^K_2. By [11, Theorem 23] $\mathrm{F}^K_3 \in \mathrm{FQC}(3, K) - \mathrm{FQC}(3, K)$. Hence the information in the table is not valid for this value of $a = 2$.

	FQC complexity	FQ complexity
$2b = a + 4$	3 or 4	2
$2b = a + 3$	2	2
$2b = a + 2$	1	1
$b = a - c$	$a - c \log c + \Theta(1)$	$\log a - \log c + \Theta(1)$
$b = a - 1$	$a - \log a + \Theta(1)$	$\log a + \Theta(1)$

6 Complexity Theory

Several of our results have analogues in complexity theory.

Definition 6.1 Let $X \subseteq \Sigma^*$ and let $k \in \mathcal{N}$. Then $\mathrm{PF}^{X[k]}$ is the set of functions that can be computed in polynomial time with k queries to X. A set $A \subseteq \Sigma^*$ is *p-superterse* if $(\forall k)(\forall X)[\mathrm{F}^A_k \notin \mathrm{PF}^{X[k-1]}]$. A function f is *k-enumerable in polynomial time* if there exists $g \in \mathrm{PF}$ such that $g(x)$ produces k values, one of which is $f(x)$. We denote this by $f \in \mathrm{EN}(k)$. Note that in this context "strongly k-enumerable" is the same as k-enumerable.

It is easy to see that analogues of Theorems 5.6 and 5.17 hold in a polynomial framework. Applying the analogue of Theorem 5.17 directly is hard since few sets have been shown to be p-superterse outright. However the following is known [1, 6, 20].

Fact 6.2 *If* $\mathrm{P} \neq \mathrm{NP}$ *then* SAT *is p-superterse.*

Combining Fact 6.2 with the polynomial analogue of Theorem 5.6 yields the following theorem.

Theorem 6.3 *Assume* $1 \le b \le a$.

1. $freq_{b,a}^{\mathrm{SAT}} \cap \mathrm{EN}(k(a, a-b)) \neq \emptyset$.
2. *If* $\mathrm{P} \neq \mathrm{NP}$ *then* $freq_{b,a}^{\mathrm{SAT}} \cap \mathrm{EN}(k(a, a-b) - 1) = \emptyset$.

7 Acknowledgments

We would like to thank Richard Chang, James Foster, Martin Kummer, Georgia Martin, Nick Reingold, Dan Spielman, and Frank Stephan for proof-reading and helpful comments.

References

[1] M. Agrawal and V. Arvind. Polynomial time truth-table reductions to p-selective sets. In *Proc. Structure in Complexity Theory ninth annual conference*. IEEE Computer Society Press, 1994.

[2] A. Amir, R. Beigel, and W. I. Gasarch. Some connections between bounded query classes and non-uniform complexity. In *Proc. of the 5th Annual Conference on Structure in Complexity Theory*, pages 232–243. IEEE Computer Society Press, July 1990. A much expanded version has been submitted to *Information and Computation*.

[3] R. Beigel. *Query-Limited Reducibilities*. PhD thesis, Stanford University, 1987. Also available as Report No. STAN-CS-88-1221.

[4] R. Beigel. A structural theorem that depends quantitatively on the complexity of SAT. In *Proc. of the 2nd Annual Conference on Structure in Complexity Theory*, pages 28–32. IEEE Computer Society Press, June 1987.

[5] R. Beigel, W. I. Gasarch, J. T. Gill, and J. C. Owings. Terse, superterse, and verbose sets. *Information and Computation*, 103:68–85, 1993.

[6] R. Beigel, M. Kummer, and F. Stephan. Approximable sets. In *Proc. of the 9th Annual Conference on Structure in Complexity Theory*, pages 12–23. IEEE Computer Society Press, 1994.

[7] J. Cai and L. A. Hemachandra. Enumerative counting is hard. *Information and Computation*, 82(1):34–44, July 1989.

[8] G. Cohen and P. Frankl. Good coverings of Hamming spaces with spheres. *Discrete Mathematics*, 56:125–131, 1989.

[9] G. Cohen, M. Karpovsky, and H. Mattson. Covering radius — survey and recent results. *IEEE Trans. Inform. Theory*, IT-31:338–343, 1985.

[10] G. Cohen, A. Lobstein, and N. Sloane. Further results on the covering radius of codes. *IEEE Trans. Inform. Theory*, IT-32:680–694, 1986.

[11] W. Gasarch. Bounded queries in recursion theory: A survey. In *Proc. of the 6th Annual Conference on Structure in Complexity Theory*, pages 62–78. IEEE Computer Society Press, June 1991.

[12] W. Gasarch, M. W. Krentel, and K. Rappoport. OptP-completeness as the normal behavior of NP-complete problems. *Mathematical Systems Theory*. to appear.

[13] V. Harizanov, M. Kummer, and J. Owings. Frequency computations and the cardinality theorem. *Journal of Symbolic Logic*, 57(2):682–687, June 1992.

[14] I. Honkala. Modified bounds for covering codes. *IEEE Trans. Inform. Theory*, IT-37:351–365, 1991.

[15] C. G. Jockusch. Semirecursive sets and positive reducibility. *Transactions of the AMS*, 131:420–436, May 1968.

[16] M. W. Krentel. The complexity of optimization problems. *J. Comput. Syst. Sci.*, 36(3):490–509, 1988.

[17] M. Kummer. A proof of Beigel's cardinality conjecture. *Journal of Symbolic Logic*, 57(2):677–681, June 1992.

[18] M. Kummer and F. Stephan. Some aspects of frequency computation, October 1991. Universität Karlsruhe, Fakultät für Informatik, Interner Bericht 21/91.

[19] M. Kummer and F. Stephan. Effective search problems. *Mathematical Logic Quarterly*, 40, 1994.

[20] M. Ogiwara. Polynomial-time membership comparable sets. In *Proc. Structure in Complexity Theory ninth annual conference*. IEEE Computer Society Press, 1994.

[21] G. Rose. An extended notion of computability. In *Abstr. International Congress for Logic, Methodology, and Philosophy of Science*, page 14, 1960.

[22] B. A. Trakhtenbrot. On the frequency computation of functions. *Algebra i Logika*, 2:25–32, 1964. (in Russian).

[23] G. V. Wee. Improved sphere bounds on the covering radius of codes. *IEEE Trans. Inform. Theory*, IT-34:237–245, 1988.