Round-query tradeoff for Single Element Recovery on affine sets

Benjamin Rossman
Duke University
September 11, 2021

Abstract
Let $F$ be a field and $A$ an affine subset of $F^N$ that does not contain the origin. How many linear queries $F^N \to F$ are required to find a nonzero coordinate of a hidden vector in $A$? We show that the deterministic $r$-round query complexity of this problem is between $r(m(A)^{1/r} - 1)$ and $rm(A)^{1/r}$ where

$$m(A) := \min_{y \in F^N : \forall x \in A, \langle x, y \rangle = 1} |\text{Supp}(y)|.$$ 

(In particular, $m(A) = N$ when $A = \{x \in F^N : \sum_i x_i = 1\}$.) The proof is a direct generalization of the special case for $F_2$ given in [Ros17].

1 Introduction
Let $\Gamma$ be a set with a zero element 0 (for example: the set $\{0, 1\}$, any field $F$, or the nonnegative real numbers $\mathbb{R}_{\geq 0}$). Let $N$ be a positive integer, and let $[N] := \{1, \ldots, N\}$.

For a vector $x = (x_1, \ldots, x_N) \in \Gamma^N$, its support is the set $\{i \in [N] : x_i \neq 0\}$ of nonzero coordinates. We denote by $\vec{0}$ the all-zero vector in $\Gamma^N$ with empty support.

For a set $D \subseteq \Gamma^N$ and a set $Q$ of functions whose domain includes $D$, we consider the following Single Element Recovery problem: For an arbitrary hidden vector $x \in D \setminus \{\vec{0}\}$, find an element of Supp($x$) (i.e., a nonzero coordinate of $x$) by making queries $q_1, \ldots, q_k \in Q$ and learning answers $q_1(x), \ldots, q_k(x)$. We consider the query complexity of this problem (i.e., minimum required number of queries) in different settings: randomized vs. deterministic, as well as non-adaptive vs. $r$-round adaptive.

For various settings of $D$ and $Q$ that have been studied, the following upper bounds are known to hold:

**Randomized non-adaptive** $O(\log^2 N)$ query algorithm. By the Valiant-Vazirani Isolation Lemma [VV85], there is a joint distribution of $t = O(\log N \cdot \log \frac{1}{\varepsilon})$ random sets $I_1, \ldots, I_t \subseteq [N]$ such that, for every nonempty $S \subseteq [N]$,

$$\Pr_{I_1, \ldots, I_t} \left[ \exists j \in \{1, \ldots, t\}, |S \cap I_j| = 1 \right] \geq 1 - \varepsilon.$$

For each $j \in \{1, \ldots, t\}$, the algorithm makes $O(\log N)$ queries that indicate whether Supp($x$) $\cap I_j$ is a singleton and, if so, give the unique coordinate in this set.

**Deterministic $r$-round** $rN^{1/r}$ query algorithm. The algorithm maintains a nonempty set of coordinates, initially $[N]$ itself, that is known to contain a nonzero coordinate of $x$. In each round, the algorithm reduces the set to one of $[N^{1/r}]$ subsets in a balanced partition. (When $r = \lfloor \log_2 N \rfloor$, this is the usual binary search using $r$ queries.)
The following table lists settings of $\mathcal{D}$ and $\mathcal{Q}$ where both upper bounds hold. The righthand columns give references for matching lower bounds. Question marks indicate the lower bound is unknown (or the author is unaware of a reference).

<table>
<thead>
<tr>
<th>$\mathcal{D}$</th>
<th>$\mathcal{Q}$</th>
<th>rand. non-adaptive $\Omega(\log^2 N)$ l.b.</th>
<th>deterministic $r$-round $\Omega(r(N^{1/r} - 1))$ l.b.</th>
</tr>
</thead>
<tbody>
<tr>
<td>${x \in {0,1}^N :</td>
<td>x</td>
<td>\text{ is odd}}$</td>
<td>XOR$^{(a)}$</td>
</tr>
<tr>
<td>${0,1}^N$</td>
<td>OR$^{(b)}$</td>
<td>[JST11]</td>
<td>[ACK20]</td>
</tr>
<tr>
<td>${ x \in {0,1}^{2\log N} : x \text{ is the char. vector of an affine subset of } \mathbb{F}_2^N }$</td>
<td>monotone$^{(c)}$</td>
<td>[KRW17]</td>
<td>?</td>
</tr>
<tr>
<td>$\mathbb{R}_{\geq 0}^N$</td>
<td>linear$^{(d)}$</td>
<td>[JST11]</td>
<td>[ACK20]</td>
</tr>
<tr>
<td>${ x \in \mathbb{F}^N : \sum_i x_i \neq 0 }$, any field $\mathbb{F}$</td>
<td>linear$^{(e)}$</td>
<td>?</td>
<td>this note</td>
</tr>
<tr>
<td>${ x \in (\mathbb{Z}/q\mathbb{Z})^N : \forall i, x_i \in {0,1} \text{ and } \sum_i x_i \neq 0 }$, any $q \geq 2$</td>
<td>linear$^{(f)}$</td>
<td>?</td>
<td>[CS21]</td>
</tr>
</tbody>
</table>

The diagram below shows implications between lower bounds from different rows of this table:

(e) $\implies$ (c) $\implies$ (d) $\implies$ (b) $\implies$ (g) $\implies$ (a) $=$ (g)$_{F_2}$ $=$ (h)$_{q=2}$ $\implies$ (h)

Most of these implications follow trivially from containments among the different classes $\mathcal{D}$ and $\mathcal{Q}$: we get a stronger lower bound whenever we are able to shrink $\mathcal{D}$ or augment $\mathcal{Q}$. Implication (f) $\implies$ (b) follows from the observation that an OR query $\bigvee_{i \in I} x_i (I \subseteq [N])$ on domain $\{0,1\}^N$ carries less information than the corresponding linear query $\sum_{i \in I} x_i$. Note that the lower bounds in rows (e), (f) and (g) imply all other lower bounds in the above table.

Some comments on specific rows in the table:

**Row (a).** Any deterministic $r$-round $q$-query algorithm in this setting converts to a depth $r + 1$, size $2^{r+q}$ AC$^0$ formula solving $\text{PARITY}_N$. Lower bounds $\Omega(N^{1/r})$ and $\Omega(r(N^{1/r} - 1))$ in row (a) follow from the AC$^0$ formula lower bounds of [Ros17]. (A weaker $\Omega(N^{1/r})$ bound follows from the much earlier AC$^0$ circuit lower bound of Håstad [Has86].) Ros17 subsequently strengthened the lower bound to $r(N^{1/r} - 1)$ (improving the previous big-$\Omega$ constant to the optimal value 1) by a linear-algebraic argument that avoids the random restriction based method of circuit complexity. (A different proof of the same $r(N^{1/r} - 1)$ lower bound appears in [CS21].) The lower bound for row (g) presented in this note is a direct generalization of the argument in [Ros17] from $F_2$ to arbitrary fields.

**Rows (d) and (e).** Here we assume that $N = 2^n$ for some integer $n$, and we identify $\{0,1\}^N$ with subsets of $\mathbb{F}_2^n$. Elements of $\mathcal{D}$ thus correspond to affine subsets of $\mathbb{F}_2^n$, and Single Element Recovery may be viewed as the search problem of finding a vector in a hidden affine set. (The lower bound in [Ros21] is presented for the slightly different, but equivalent, search problem of finding a nonzero vector in a hidden non-trivial linear subspace of $\mathbb{F}_2^n$.)
Row (b). [ACK20] show that $N^{1/r} - 1$ queries are required in at least one round. A straightforward extension of their argument yields a stronger $r(N^{1/r} - 1)$ lower bound on the total number of queries.

Row (f). [ACK20] show that $\frac{1}{2}(N^{1/r} - 1)$ linear queries $\mathbb{R}^n \to \mathbb{R}$ are required in at least one round. It is again a straightforward extension to obtain a stronger $\frac{1}{2}r(N^{1/r} - 1)$ lower bound on the total number of queries. Here the factor $\frac{1}{2}$ arises from splitting each linear query into two nonnegative linear queries; it is unclear if this factor is necessary for domain $\mathbb{R}^N$. (If so, the implication (f) $\Rightarrow$ (g)$_R$ holds up to a factor 2.) The $r(N^{1/r} - 1)$ lower bound for row (g), proved in this notes, shows that this factor is unnecessary for domain $\mathbb{R}^N$.

Rows (g) and (h). When $q$ is prime, we may compare the $r$-round lower bounds of rows (g) and (h) for the field $\mathbb{F}_q = \mathbb{Z}/q\mathbb{Z}$. For $q = 2$, both (g) and (h) specialize to the $rN^{1/r} - 1$ bound of row (a). However, these bounds are incomparable for primes $q \geq 3$. Quantitatively, the $r$-round lower bound in row (g) is $rN^{1/r} - 1$ for all fields $\mathbb{F}$, while row (h) is $\Omega(\frac{r(N^{1/r} - 1)}{q^{1+1/r} \log^2 q})$; for this reason, (h) $\not\Rightarrow$ (g). On the other hand, the domain of (h) is smaller than the domain of (g), as it includes the additional constraint that $x \in \{0, 1\}^N$; for this reason, (g) $\not\Rightarrow$ (h).

Definition 1. If $A$ is an affine subset of $\mathbb{F}^N$ that does not contain the origin, we define $m(A) = 1$ by

$$m(A) := \min_{y \in \mathbb{F}^N : \forall x \in A,\, (x,y)^T = 1} |y|.$$ 

For example, if $A$ is the codimension-1 affine set $\{x \in \mathbb{F}^N : \sum x_i = 1\}$, then we have $m(A) = N$.

Theorem 2. Let $A$ be an affine subspace of $\mathbb{F}^N$ that does not contain the origin. Then the number of linear queries $\mathbb{F}^N \to \mathbb{F}$ required to solve Single Element Recovery on domain $A$ is at least $r(m(A)^{1/r} - 1)$ and at most $rm(A)^{1/r}$.

2 Round-query tradeoff over any field

Fix an arbitrary field $\mathbb{F}$ and positive integer $N$. We consider the vector space $\mathbb{F}^N$. For $x \in \mathbb{F}^N$, let $|x|$ denote the number of nonzero coordinates of $x$ (i.e., the size of support of $x$, not the sum of coordinates $\sum_i x_i$).

In what follows, $U, V, S, T$ represent linear subspaces of $\mathbb{F}^N$, while $A, B$ represent affine subsets of $\mathbb{F}^N$ that do not contain the origin. Notation $U <_k V$ expresses that $U$ is a codimension-$k$ subspace of $V$.

Fix an inner product $\langle\cdot, \cdot\rangle$ on $\mathbb{F}^N$. Let $V^\perp$ denote the orthogonal complement of a linear subspace $V$, that is, $V^\perp := \{x \in \mathbb{F}^N : \langle x, v \rangle = 0 \text{ for all } v \in V\}$. Note that $(V^\perp)^\perp = V$ and $U <_k V \Rightarrow V^\perp <_k U^\perp$.

Lemma 3. Let $A$ be an affine subset of $\mathbb{F}^N$ that does not contain the origin. Let $U <_1 V$ be linear subspaces

$$U := \{x_1 - x_2 : x_1, x_2 \in A\} \quad \text{and} \quad V := \{\lambda x : x \in A, \lambda \in \mathbb{F}\}.$$ 

Then $m(A) = \min_{z \in U \setminus V^\perp} |z|$.

Lemma 4. For all $U <_k V \leq \mathbb{F}^N$, there exists a linear projection $\pi : V \to U$ such that $|\pi(v)| \leq (k + 1)|v|$ for all $v \in V$.

Proof. Let $w_1, \ldots, w_k$ be a basis of $V$ over $U$, greedily chosen so that $|w_i|$ is maximal among elements $w_i \in V \setminus \text{LinearSpan}(U \cup \{w_1, \ldots, w_{i-1}\})$ for each $i \in \{1, \ldots, k\}$. Consider any $v \in V$. Then $v = u + a_1 w_1 + \cdots + a_k w_k$ for some $u \in U$ and $a_1, \ldots, a_k \in \mathbb{F}$. Let $\pi : V \to U$ be the projection $v \mapsto u$. Then

$$|\pi(v)| = |u| = |v - (a_1 w_1 + \cdots + a_k w_k)| \leq |v| + |a_1 w_1| + \cdots + |a_k w_k|.$$ 

For each $i$, we either have $a_i = 0$, in which case $|a_i w_i| = 0$; or $a_i \neq 0$, in which case $v \notin \text{LinearSpan}(U \cup \{w_1, \ldots, w_{i-1}\})$ and therefore $|v| \leq |w_i| = |a_i w_i|$. We conclude that $|\pi(v)| \leq (k + 1)|v|$ as required. 

\qed
Definition 5. For real \( r, m \geq 0 \), let
\[
\beta(r, m) := \lim_{\rho \to r} \rho(m^{1/\rho} - 1) = \begin{cases} 
0 & \text{if } r = 0 \text{ and } m \leq 1, \\
\infty & \text{if } r = 0 \text{ and } m > 1, \\
r(m^{1/r} - 1) & \text{if } r > 0.
\end{cases}
\]

As an aside: note that \( \lim_{r \to \infty} \beta(r, m) = \ln(m) \).

Lemma 6. For all real \( r, m, k \geq 0 \),
\[
\beta\left(\frac{r}{m} \ell\right) + \ell - 1 \geq \beta(r + 1, m).
\]
When \( r > 0 \), this holds with equality if and only if \( \ell = m^{1/(r+1)} \).

Proof. Elementary calculus.

Theorem 7 (Lower Bound). Let \( A \) be an affine subspace of \( \mathbb{F}^N \) that does not contain the origin. Then \( r \)-round Single Element Recovery on domain \( A \) requires at least \( \beta(r, m(A)) \) linear queries.

Proof. We prove the theorem by induction on \( r \). In the base case \( r = 0 \), the theorem states that Single Element Recovery on domain \( A \) is solved in zero rounds (with zero queries) if, and only if, \( m(A) = 1 \). This follows from the definition of \( m(A) \).

For the induction step, let \( r \geq 1 \) and assume the theorem holds for fewer than \( r \) rounds. Suppose an optimal deterministic \( r \)-round algorithm makes \( k \) queries in the first round. These queries correspond to a linear function \( \varphi : \mathbb{F}^N \to \mathbb{F}^k \).

For each possible sequence of answers \( b \in \mathbb{F}^k \), we are left with a deterministic \( r-1 \)-round algorithm that solves Single Element Recovery over \( \{ x \in A : \varphi(x) = b \} \), which is a codimension-\( k \) affine subset of \( A \). We show how to adversarially choose an answer sequence \( b \in \mathbb{F}^k \) such that the remaining \( r-1 \)-round algorithms requires at least \( \beta(r-1, m(A)/(k+1)) \) queries. The total number of queries is then at least \( k + \beta(r-1, m(A)/(k+1)) \), which is \( \geq \beta(r, m(A)) \) by Lemma 6.

We proceed to explain how to find a suitable answer sequence \( b \in \mathbb{F}^k \) for linear queries \( \varphi : \mathbb{F}^N \to \mathbb{F}^k \). As in Lemma 3, let \( U <_1 V \) be linear subspaces
\[
U := \{ x_1 - x_2 : x_1, x_2 \in A \} \quad \text{and} \quad V := \{ \lambda x : x \in A, \lambda \in \mathbb{F} \}.
\]
Let
\[
S := \text{Ker}(\varphi) \cap U.
\]
Without loss of generality, we may assume that \( S <_k U \), since otherwise a proper subset of the \( k \) queries would contain the same information about a hidden vector in \( A \).

By duality, we have \( V^\perp <_1 U^\perp <_k S^\perp \). By Lemma 4 there exists a linear projection
\[
\pi : S^\perp \to U^\perp
\]
such that \( |\pi(z)| \leq (k+1)|z| \) for all \( z \in S^\perp \). Let
\[
T := (\pi^{-1}(V^\perp))^\perp.
\]
By duality and the rank-nullity theorem, we have

\[ T^\perp = \pi^{-1}(V^\perp), \quad V^\perp \leq_k T^\perp <_k S^\perp, \quad S \leq_k T <_k V, \quad S = T \cap U. \]

Note that \( T \not\subseteq U \), since otherwise \( S = T \) (whereas we have shown \( S <_1 T \)). Fix any \( t \in T \setminus U \). Since \( V = \{ \lambda x : x \in A, \lambda \in \mathbb{F} \} \), there exist \( x_0 \in A \) and \( c \in \mathbb{F} \) such that \( cx_0 = t \); moreover, \( c \neq 0 \) (since \( t \neq 0 \) as \( t \notin U \)). Now let

\[ B := \{ s + c^{-1}t : s \in S \}. \]

We claim that

\[ B = \{ x \in A : \varphi(x) = \varphi(t) \}. \]

We first prove the containment \( \subseteq \). Consider any element \( s + c^{-1}t \in B \) where \( s \in S \). Since \( c^{-1}t = x_0 \in A \) and \( S \subseteq U = \{ x_1 - x_2 : x_1, x_2 \in A \} \), we have \( s + c^{-1}t = x_0 + x_1 - x_2 \) where \( x_0, x_1, x_2 \in A \). Therefore, \( s + c^{-1}t \in A \). Having shown \( B \subseteq \{ x \in A : \varphi(x) = \varphi(t) \} \), it then follows that these sets are equal, since both are affine sets of dimension \( \text{dim}(A) - k \).

Note that

\[ S = \{ w_1 - w_2 : w_1, w_2 \in B \} \quad \text{and} \quad T = \{ \lambda w : w \in B, \lambda \in \mathbb{F} \}. \]

We have

\[
m(B) = \min_{z \in S^\perp \setminus T^\perp} |z| = \min_{z \in \pi^{-1}(V^\perp) \setminus \pi^{-1}(V^\perp)} |z| \geq \min_{z \in \pi^{-1}(V^\perp) \setminus \pi^{-1}(V^\perp)} \frac{|\pi(z)|}{k + 1} = \min_{y \in U^\perp \setminus V^\perp} \frac{|y|}{k + 1} = \frac{m(A)}{k + 1}.
\]

Let \( \varphi(t) \in \mathbb{F}^k \) be the (adversarially chosen) sequence of answers to first-round queries \( \varphi : \mathbb{F}^N \to \mathbb{F}^k \). We are left with an \( r - 1 \) round algorithm that solves Single Element Recovery on domain \( B \). By the induction hypothesis and Lemma 6, the total number of queries over all \( r \) rounds is at least

\[
k + \beta(r - 1, m(B)) \geq k + \beta(r - 1, \frac{m(A)}{k + 1}) \geq \beta(r, m(A)),
\]

which completes the proof. \( \square \)

**Theorem 8 (Upper Bound).** Let \( A \) be an affine subspace of \( \mathbb{F}^N \) that does not contain the origin. Then for all \( r \geq 1 \), Single Element Recovery on domain \( A \) is solvable in \( r \) rounds using \( rm(A)^{1/r} \) linear queries.

**Proof.** Fix \( y \in \mathbb{F}^N \) such that \( |y| = m(A) \) and \( \langle x, y \rangle = 1 \) for all \( x \in A \). Let \( I = \{ i \in [N] : y_i \neq 0 \} \) (so, \( |y| = |I| = m(A) \)). Let \( I = I_1 \sqcup \cdots \sqcup I_k \) be partition of \( I \) into \( k \leq [m(A)^{1/r}] \) blocks of size at most \( m(A)^{(r-1)/r} \). For \( j \in \{1, \ldots, k\} \), let \( y^{(j)} \) be the sub-vector of \( y \) with support \( I_j \).

In the first round, the algorithm makes \( k - 1 \) queries that learn the values of \( \langle x, y^{(1)} \rangle, \ldots, \langle x, y^{(k-1)} \rangle \). From the answers, we learn some \( j \in \{1, \ldots, k\} \) and \( \lambda \in \mathbb{F}^\times \) such that \( \langle x, y^{(j)} \rangle = \lambda \). The algorithm then proceeds
to solve Single Element Recovery on the corresponding affine subset $B$ of $A$. We have $m(B) \leq m(A)^{(r-1)/r}$, as witnessed by the fact that $\langle w, \lambda^{-1} y^{(j)} \rangle = 1$ for all $w \in B$.

If $r = 1$, we are done, since $|y^{(j)}| = m(A)^0 = 1$ and we have used $k - 1 = m(A) - 1$ queries. If $r \geq 2$, then by induction the total number of queries is at most

$$k - 1 + (r - 1)m(B)^{1/(r-1)} \leq \lceil m(A)^{1/r} \rceil - 1 + (r - 1)(m(A)^{(r-1)/r})^{1/(r-1)} \leq m(A)^{1/r} + (r - 1)m(A)^{1/r} = rm(A)^{1/r}.$$ 

\[ \square \]

References


