Bit-probe lower bounds for succinct data structures

Emanuele Viola

Northeastern University

June 2009
Bits vs. trits

- Store $n$ “trits” $t_1, t_2, \ldots, t_n \in \{0, 1, 2\}$

- In $u$ bits $b_1, b_2, \ldots, b_u \in \{0, 1\}$

- Want:
  - Small space $u$ (optimal $= \lceil n \log_2 3 \rceil$)
  - Fast retrieval: Get $t_i$ by probing few bits (optimal $= 2$)
Two solutions

- Arithmetic coding:
  Store bits of \((t_1, \ldots, t_n) \in \{0, 1, \ldots, 3^n - 1\}\)

  Optimal space: \(\lceil n \log_2 3 \rceil \approx n \cdot 1.584\)

  Bad retrieval: To get \(t_i\) probe all > \(n\) bits

- Two bits per trit

  Bad space: \(n \cdot 2\)

  Optimal retrieval: Probe 2 bits
Divide $n$ trits $t_1, \ldots, t_n \in \{0,1,2\}$ in blocks of $q$.

Arithmetic-code each block.

Space: $\left\lceil q \log_2 3 \right\rceil \frac{n}{q} < (q \log_2 3 + 1) \frac{n}{q}$

= $n \log_2 3 + \frac{n}{q}$

Retrieval: Probe $O(q)$ bits

Polynomial tradeoff between redundancy, probes.
Polynomial tradeoff

- Divide \( n \) trits \( t_1, \ldots, t_n \in \{0,1,2\} \) in blocks of \( q \)
- Arithmetic-code each block

Space: \( \lceil q \log_2 3 \rceil \frac{n}{q} = (q \log_2 3 + 1/q^{\Theta(1)}) \frac{n}{q} = n \log_2 3 + n/q^{\Theta(1)} \)

Retrieval: Probe \( O(q) \) bits
Exponential trade-off

- Breakthrough [Pătraşcu '08, later + Thorup]

Space: $n \lg_2 3 + n/2^{\Omega(q)}$

Retrieval: Probe $q$ bits

- E.g., optimal space $\lceil n \lg_2 3 \rceil$, probe $O(lg n)$
Our results

• **Theorem [this work]:**
  Store $n$ trits $t_1, \ldots, t_n \in \{0,1,2\}$ in $u$ bits $b_1, \ldots, b_u \in \{0,1\}$.
  If get $t_i$ by probing $q$ bits then space $u > n \lg 3 + n/2^{\Omega(q)}$.

• Matches [Pătraşcu Thorup]: space $< n \lg 3 + n/2^{\Omega(q)}$

• Holds even for adaptive probes
Outline

- Bits vs. trits
- Bits vs. sets
- Proof
**Bits vs. sets**

- Store $S \subseteq \{1, 2, \ldots, n\}$ of size $|S| = k$

  $\begin{array}{c}
  \text{In } u \text{ bits } b_1, \ldots, b_u \in \{0,1\} \\
  \end{array}$

- Want:
  - Small space $u$ (optimal $= \lceil \lg_2 \binom{n}{k} \rceil$)
  - Answer “$i \in S$?” by probing few bits (optimal $= 1$)
Previous results

- Store \( S \subseteq \{1, 2, \ldots, n\} \), \(|S| = k\) in bits, answer “\( i \in S? \)”

- [Minsky Papert '69] Average-case study

- [Buhrman Miltersen Radhakrishnan Venkatesh; Pagh '00] Space \( O(\text{optimal}) \), probe \( O(\lg(n/k)) \)

  Lower bounds for \( k < n^{1-\varepsilon} \)

- No lower bound was known for \( k = \Omega(n) \)
Theorem [this work]:
Store \( S \subseteq \{1, 2, \ldots, n\} \), \(|S| = n/3\) in \( u \) bits \( b_1, \ldots, b_u \in \{0,1\} \).

If answer “\( i \in S? \)” probing \( q \) bits then space \( u > \text{optimal} + \frac{n}{2^{\Omega(q)}} \).

- First lower bound for \(|S| = \Omega(n)\)
- Holds even for adaptive probes
Outline

• Bits vs. trits

• Bits vs. sets

• Proof
Recall our results

• Theorem:
  Store $n$ trits $t_1$, ..., $t_n \in \{0,1,2\}$
  in $u$ bits $b_1$, ..., $b_u \in \{0,1\}$.

  If get $t_i$ by probing $q$ bits
  then space $u > n \log_2 3 + n/2^O(q)$.

• For now, assume non-adaptive probes:
  $t_i = d_i (b_{i1}, b_{i2}, ..., b_{iq})$
Proof idea

- \( t_i = d_i (b_{i1}, b_{i2}, \ldots, b_{iq}) \)

- Uniform \( (t_1, \ldots, t_n) \in \{0,1,2\}^n \)
  
  Let \( (b_1, \ldots, b_u) := \text{Store}(t_1, \ldots, t_n) \)

- Space \( u \approx \text{optimal} \Rightarrow (b_1, \ldots, b_u) \in \{0,1\}^u \approx \text{uniform} \Rightarrow \)

  \[
  1/3 = \Pr [ t_i = 2 ] = \Pr [ d_i (b_{i1}, \ldots, b_{iq}) = 2 ] \approx A / 2^q \neq 1/3
  \]

Contradiction, so space \( u >> \text{optimal} \)

Q.e.d.
Handling adaptivity

- So far $t_i = d_i(b_{i1}, b_{i2}, \ldots, b_{iq})$

- In general, $q$ adaptively chosen probes = decision tree
  
  $2^q$ bits
  depth $q$

$1/3 = \Pr[t_i = 2] = \Pr[d_i(b_{i1}, \ldots, b_{iq}) = 2] \approx A/2^q \neq 1/3$
Remarks on proof

- Use ideas from lower bounds for locally decodable codes [Shaltiel V.]

- New approach to data structures lower bounds
Conclusion

• Thm: Store n trits $t_1, \ldots, t_n \in \{0,1,2\}$.
  Get $t_i$ by probing $q$ bits $\Rightarrow$ space $> \text{optimal} + \frac{n}{2^{O(q)}}$

  Matches [Pătraşcu Thorup]: space $< \text{optimal} + \frac{n}{2^{\Omega(q)}}$

• Thm: Store $S \subseteq \{1, 2, \ldots, n\}$, $|S| = n/3$.
  Answer “$i \in S$?” probing $q$ bits $\Rightarrow$ space $> \text{optimal} + \frac{n}{2^{O(q)}}$

  First lower bound for $|S| = \Omega(n)$

• New approach to lower bounds for basic data structures