Locality in Coding Theory

Madhu Sudan
Harvard University
Locality ≠ LOCALity

- Local ≡ “Influenced by small neighborhood”
- Two views in literature:
  - Focus on “small”
    - Leads to greater efficiency in algorithms, but not necessarily structure.
  - Focus on “neighborhood”
    - Suitable in distributed computing.
    - More structure in algorithms
    - Input to algorithm = architecture for solution
- Coding Theory: Focus on “small”
Locality in Coding Theory

- Historical accident?
- Classical concepts:
  - Local decodability
  - Local list-decodability
  - Local testability
- Striking recent developments:
- Modern Concepts:
  - Local Reconstruction Codes
  - Some constructions
(Classical) Challenge of Coding Theory

- Protect data from (inevitable) error
- Question: Does ability to protect go \( \uparrow \) with size \( \uparrow \)
  - Shannon’s surprising answer (surprise in 1948): YES
  - Cost = ? Recovery time \( \uparrow \) with size \( \uparrow \)
- Current options to store massive data:
  - Break data into small chunks.
    - Good recovery time but lose twice in reliability: (1) Reliability per chunk; (2) Union bound on chunks.
  - Keep data in one big chunk:
    - Good reliability, but poor recovery time.
- Locality (if achievable): Best of both worlds.
  - Recovery time = sublinear?
Basic Notions:

- **Code:** $E: \Sigma^k \rightarrow \Sigma^n$; $\delta(E) = \min_{x \neq y} \{\delta(E(x), E(y))\}$
  - Testing: Given $E, r \in \Sigma^n$; $\exists m$ s.t. $r = E(m)$?
  - Decoding: Given $E, r$ find $m$ s.t. $\delta(r, E(m)) \leq \delta_0$

- **Locality:** Do above in sublinear time!
  - Local testing:
    - Make few ($\ell$) queries to $r$;
    - Reject with prob. $\geq \epsilon \cdot \delta(r, E)$

- **Locally Decodable (correctible) Code:**
  - Given $E, r, i$ make few ($\ell$) queries to $r$; output $E(m)_i$

- **Local Reconstruction Codes:**
  - LDCs for $\delta_0 = \frac{1}{n}$; (even simpler: erasures only!)
Progress Report: LTCs

- **Motivations:**
  - Entirely mathematical?
    - PCPs ... XOR lemma, extremal graphs, ...

- **Initial belief: Locality** $c(\epsilon) \to \infty$ as $n = k^{1+\epsilon}; \epsilon \to 0$

- **Disproved many times:**
  - [GS,BGHSV,BS,Dinur,Meir]: $c = 3$ with $n = k \cdot \text{polylog } k$

- **Semi-recent belief: Locality** $c = k^\epsilon \Rightarrow \frac{k}{n} \to 0$ as $\epsilon \to 0$

- **Disproved fewer times:**
  - [Viderman’11]: $c = k^\epsilon$ and $\frac{k}{n} = 1 - \delta$, $\forall \epsilon, \delta > 0$.
  - [Kopparty,Meir,Ron-Zewi,Saraf’16]: $c = \text{polylog}(k)$ achievable for “free”

- **Practical use?**
Progress Report: LDCs

- **Motivations:**
  - Mathematical ... + (in hindsight) Practical?

- **Initial belief: Locality** \( \ell(\varepsilon) \to \infty \) as \( n = 2^{k\varepsilon}; \varepsilon \to 0 \)
  - Disproved (2006-9) [Yekhanin, Raghavendra, Efremenko]: \( \ell = 3; n = \exp\left(\exp\left(\sqrt{\log k}\right)\right) \)

- **Semi-recent belief: Locality** \( \ell = k^\varepsilon \Rightarrow \frac{k}{n} \to 0 \) as \( \varepsilon \to 0 \)
  - Disproved “many times” (2010-2016):
    - [Kopparty, Meir, Ron-Zewi, Saraf’16]: \( \ell = 2^{\sqrt{\log k}} \) achievable for “free”

- **Practical Use?!**
Local Reconstruction Codes

- Practically motivated notion:
  - Fix small corruption quickly (locally).
  - Remain resilient to many errors.

- $(\ell, d)$-LRC [Gopalan-Huang-Simitci-Yekhanin]:
  - Code of distance $d/n$ s.t. single symbol erasure can be recovered $\ell$-locally.
    - Symbol of message? (Weak Defn.)
    - Symbol of codeword? (Strong Defn.)
An Easy Construction

\[ n = k + \frac{k}{\ell} + d - 1 \]

Satisfies weak defn.

Can save \( \oplus \) if it is \( \oplus \) of all
A limiting result

- Recall Singleton Bound: \( n \geq k + d - 1 \)
  - Project Code to first \( k - 1 \) coordinates.
  - Pigeonhole Principle \( \Rightarrow \) 
    \[ \exists x \neq y \text{ s.t. } E(x)_{1:k-1} = E(y)_{1:k-1} \]
- Simple \((\ell, d)\)-LRC weaker by \( \approx + \frac{k}{\ell} \)
- Thm [GHSY]: This is necessary (up to \( \pm 1 \))
  - Find \( k - 1 + (k - 1)/\ell \) locations of rank \( k - 1 \) greedily.
  - Project to these coordinates + PHP.
A “state-of-art” construction

- [Tamo-Barg]: \( n = k + \frac{k}{\ell} + d (\pm 1) \) (Strong Defn.)

- Code = Subcode of Reed-Solomon Code:
  - Work with \( \mathbb{F}_q \) where \( r = \ell + 1 \) divides \( q; \; \omega^r = 1 \)
  - \( \{ (f(\alpha)|\alpha \in \mathbb{F}_q^*) | f(x) = \sum_{\{i \in [k-1]; i \mod r \neq -1\}} m_i x^i \} \)
  - \( (\ell = 2; p(x) = a + bx + cx^3 + dx^4 + ex^6 + \ldots) \)
  - Local groups: \( S_a = \{a, a\omega, a\omega^2, \ldots, a\omega^\ell\}; \; r \) points
  - \( f|_{S_a} \equiv f(x) \mod (x^r - a^r); \; \text{deg.} \leq r - 2! \)
  - \( (p(x) \mod (x^3 - 1) = (a + c + e \ldots) + (b + d \ldots) x) \)
Conclusions

- Locality in coding theory
  - Diverse collection of notions
  - New problems and solutions (and techniques).
  - Diverse applications.
  - Many open questions!