A Crash Course on Coding Theory

Madhu Sudan
MIT

DIMACS & IBM Workshop
IBM Almaden Research Center
6-10 November, 2000

Disclaimer

This is an opinionated survey of coding theory, unbiased by actual reading of papers.

Some Opinions

Aka: Table of Contents

What Coding Theory has to offer:

- Constructions of Error-correcting codes.
- Bounds (limitations) on the performance of error-correcting codes.
- Algorithms for error-correction.
- Connections to other fields (in our case Theory of Computation).

Some Canonical References

- The Handbook of Coding Theory, volumes I and II.
  - Has everything you want and more.
  - Very much current.
  - Some excellent chapters (e.g. applications to deep-space communcation, algebraic-geometry codes).
  - Cost = $300.
  - Sometimes a bit excessive (e.g. 130 pages of table of best known codes).
- MacWilliams and Sloane: More compressed than above, but a bit outdated.

Some Canonical References (contd.)

- van Lint: Much more handy than above.
- Richard E. Blahut: Stolen from MIT library; must be good! (Update (10/15/2000): Found the book! Is good! Highly recommended.)
- Berlekamp: "The reader looking for simple or elementary proofs is warned ..."
- "Key papers in the development of coding theory": Terrific source book!

In general, not enough emphasis on algorithms. Blahut's book is best source for algorithms. van Lint is good for quick reference.

Breakdown of lectures

- History and definitions
- Constructions - 1
- Bounds
- Algorithms
  - Classical RS + linear codes.
  - List decoding; Forney's GMD
  - Linear time algorithms.
  - Random error vs. adversarial error.
- Complexity
- Modern day things: Probabilistic error-correction?

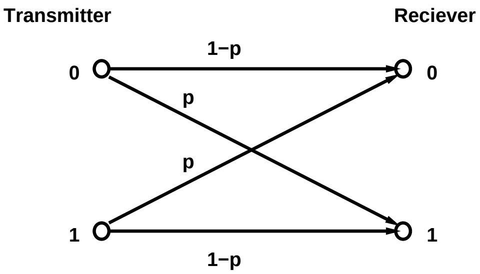
History: Ode to Shannon

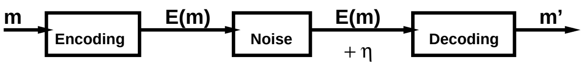
- Clearly everything started with Shannon's paper titled "A Mathematical Theory of Communication".
- Foundations of Information Theory, as well as Coding Theory. Notion of Entropy of Information.
- Two models of communication: Noiseless and Noisy.
- Goal in former: Compress information to take advantage of redundancy in data. Examples such as: Entropy of English. Coding for the Morse code etc. Leads to Noiseless Coding Theorem.

History: Ode to Shannon (contd.)

- Goal in latter: Add redundancy in data to compensate for channel noise. Leads to Noisy Coding Theorem.
- Coding theory originates from latter.

Example: Binary Symmetric Channel





E: Maps $k = Rn$ bits to n bits.
D: Maps n bits to Rn bits.
R: Rate of source < 1.

Fundamental question

$e(R, p) =$ Freq. of error as $n \to \infty$.

$$e(R, p) = \lim_{n \to \infty} \{\Pr_{\eta,m}[D(E(m)) \neq m]\}$$

Belief: $R > 0 \quad \Rightarrow \quad e(R, p) > 0$

Noisy coding theorem:
$\forall p < 1/2, \exists C(p) > 0$
s.t. if $R < C(p)$ then $e(R, p) = 0$.

Converse coding theorem:
$\forall p < 1/2, \exists$ (same) $C(p) > 0$
s.t. if $R > C(p)$ then $e(R, p) = 1$.

Some Notations

Hamming Distance: For $x, y \in \Sigma^n$,
$\Delta(x, y) = \#$ coordinates s.t. $x_i \neq y_i$.

Hamming Ball:
$B(x, r) = \{y | \Delta(x, y) \leq r\}$

Binary Entropy Function:
$H(p) = -(p \log_2 p + (1-p) \log_2 (1-p))$

Fact: Hamming ball of radius pn has approximately $2^{H(p)n}$ elements.

Addendum to Shannon's Theorem:
Capacity of Binary symmetric channel

$$C(p) = 1 - H(p)$$

.

Proof of Coding Theorem

(Uses Probabilistic Method)

Encoding: $E : \{0,1\}^{Rn} \to \{0,1\}^n$ random.

Decoding: Given y, if $\exists \, ! \, x$ such that $E(x) \in B(y, (1+\epsilon)pn)$, then $D(y) = x$, else arbitrary.

Analysis (ignoring ϵ):
$\Pr[$ Decoding Error $]$
$\leq \Pr[\ \#$ errors $> pn]$ (1)
$+ \Pr[$ diff. codeword in $B(y, pn)]$ (2)

Prob. (1) small by Chernoff Bounds.
Prob. (2) at most $2^{H(p)n} \cdot 2^{Rn} \cdot 2^{-n}$
$\leq \exp(-n)$, if $R < 1 - H(p)$

(Proof shows good E exists.)

Proof of Converse

Transmit random msg.; decoding error = ?

D partitions $\{0,1\}^n$ into $S_1, \ldots, S_K$:
S_i decoding to ith message. ($K = 2^{Rn}$)

Key observations:
(1) $\Pr[\#$ errors $\leq (1-\epsilon)pn]$ very small.
(2) If $y \notin B(E(m_i), (1-\epsilon)pn)$
$\Pr[E(m_i) + \eta = y] \leq 2^{-H((1-\epsilon)p)n}$

Analysis (ignoring ϵ):
Prob. decoding correctly
$\leq$ Prob. (1)
$+ \sum_i \Pr[\text{transmit } m_i] \cdot |S_i| \cdot 2^{-H(p)n}$
$=$ Prob. (1) $+2^{(1-H(p)-R)n}$.

Variants of Theorem

Strong form of coding theorem:
$\forall p < 1/2, \exists C(p) > 0$
s.t. if $R < C(p)$ then $e_n(R,p) = 2^{-En}$,
where $E = E_{R,p} > 0$.

E the error exponent is still a subject of investigation.

Profound form of coding theorem:
$\forall$ noisy channel, $\exists$ capacity
$\forall$ source, $\exists$ rate
s.t. if rate $<$ capacity,
then information transmission is feasible.

Algorithmic Goals

- Compute E, D in polynomial time, while minimizing $=e(R,p)$.
- ... in linear time?
- For other models of error: η is not i.i.d.?
- E, D to minimize $\mathrm{E}[\Delta(m, D(E(m)))]$.

Most questions still being studied.

Combinatorial Coding Theory

- $e(R,p)$ too hard to analyze, given E, D.
- Lets study: $\min_{m \neq m'}\{\Delta(E(m), E(m'))\}$.
- Code $\mathcal{C} = \{E(m) | m\}$, over alphabet Σ:

Defn: $\mathcal{C} \subseteq \Sigma^n$ with $|\Sigma| = q,\ |\mathcal{C}| = q^k$ called an $(n,k)_q$ code.

$$\Delta(\mathcal{C}) = \min_{\text{distinct } x,y \in \mathcal{C}} \{\Delta(x,y)\}$$

$\mathcal{C}$ with $\Delta(\mathcal{C}) = d$ also called $(n,k,d)_q$ code.

Warning: Sometimes call this an $(n, q^k, d)_q$ code!

Standard Terminology

If $\mathcal{C} = (n, k, d)_q$ code, then:

- n = Block Length.
- k = Information Length.
- d = Distance.
- k/n = (Information) Rate.
- d/n = Distance (Rate).
- q = (Alphabet size).

(Words within parenthesis often omitted.)

Aside: Finite fields

Often Σ is a field of size q.

Fact: $\forall$ prime powers q, $\exists$ field of size q.

Fact: Given prime p, integer k,
field of size $q = p^k$ can be "computed"
in time $\mathrm{poly}(p, k)$
stored in space $\mathrm{poly}(\log p, k)$
s.t. field operations can be carried out
in time $\mathrm{poly}(\log p, k)$.

Fact: $\forall\ l \geq 0$, $\exists$ explicit field of size $2^{2\cdot 3^l}$.

(Field given by irred. poly of deg. k.
$x^{2\cdot 3^l} + x^{3^l} + 1$ is irreducible over Z_2.)

Linear Codes

If Σ field, then Σ^n vector space.

If $\mathcal{C}$ linear subspace, then $\mathcal{C}$ linear code.

Denote $[n, k]_q$ or $[n, k, d]_q$ code.

Niceness of Linear Codes

Generator Matrix:
$\mathcal{C}$ linear $\Rightarrow$ $\exists$ $k \times n$ matrix G s.t.
$\mathcal{C} = \{xG | x \in \Sigma^k\}$

Parity Check Matrix:
$\mathcal{C}$ linear $\Rightarrow$ $\exists$ $n \times (n-k)$ matrix H s.t.
$\mathcal{C} = \{y \in \Sigma^n | yH = 0\}$

Implications:

- $\mathcal{C}$ can be represented succinctly.
- Encoding is efficient.
- Error-detection is efficient.
- "Syndrome" (yH) has error information.
- Gives q^{n-k} sized table for decoding.
(Useful if $n - k$ small.)

Columns of Parity check define $[n, n-k]_q$ code called the dual code $\mathcal{C}^\perp$.

Distance vs. Weight

Defn: $\mathrm{wt}(x) = \#$ non-zero coordinates of x.

Observe $\Delta(x, y) = \mathrm{wt}(x - y)$.

Thus $\Delta(\mathcal{C}) = \min_{\vec{0} \neq x \in \mathcal{C}} \{\mathrm{wt}(x)\}$.

(Note $0 \in \mathcal{C}$ for every linear code $\mathcal{C}$.)

Basic questions in Coding theory

- Given n, k, q, find $(n, k)_q$ code $\mathcal{C}$ that maximizes $\Delta(\mathcal{C})$.
- Given n, d, q, find $(n, k)_q$ code $\mathcal{C}$ with $\Delta(\mathcal{C}) \geq d$ that maximizes k.
- Given k, δ, q, find $(n, k)_q$ code $\mathcal{C}$ with $\Delta(\mathcal{C}) \geq \delta n$ that minimizes n. (Better phrasing of algorithmic question.)
- Given n, k, d, find $(n, k)_q$ code $\mathcal{C}$ with $\Delta(\mathcal{C}) \geq d$ that ???imizes q. (Actually a very nice perspective.)