Distance verification for LDPC codes

Ilya Dumer
UC Riverside, USA

Background: MLD and DistVer (finding MinDist) for linear $[n, k]$-codes.

A. MLD is NP-hard (Berlecamp-McEliece-van Tilborg 1978)

B. DistVer is NP-hard (Vardy 1991)

C. DistApprox is NP-hard within a const factor or a linear additive error
(Dumer-Micciancio-Sudan 1999 - RP reductions; Cheng-Wan 2009)

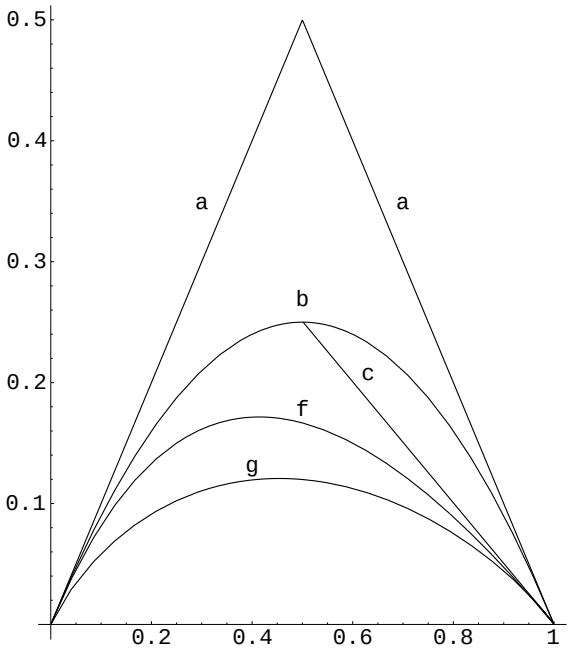
- **Algorithms**: for generic $[n, k]$-codes of rate R, DistVer and MLD require expon. complexity $2^{F(R)n}$. We discuss 3 algorithms:

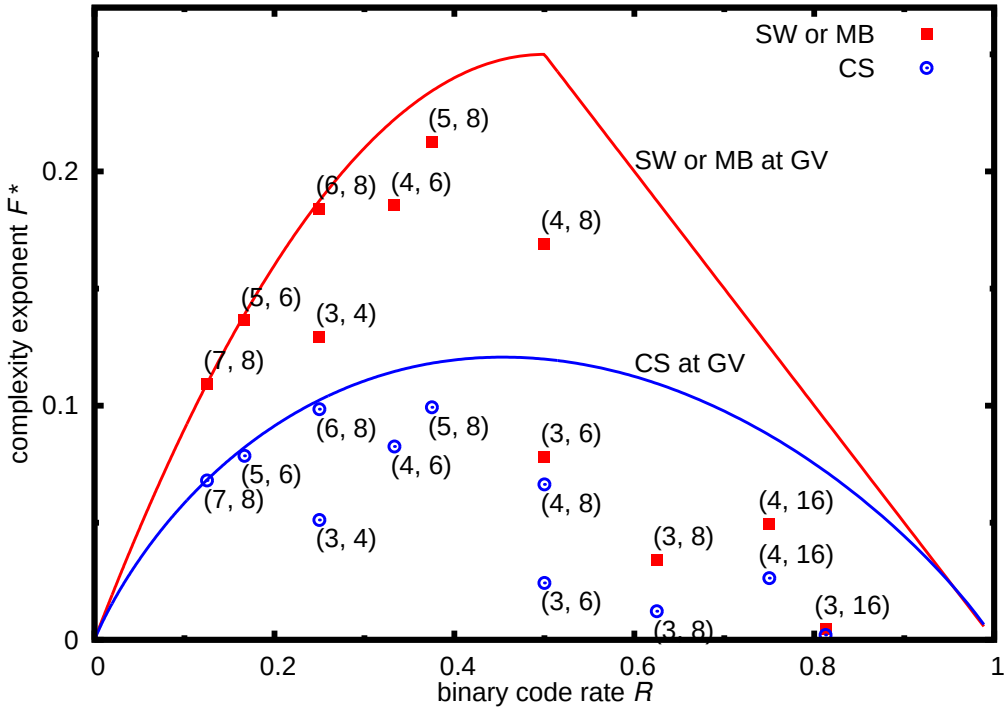
  Algorithm 1: Correct sliding k-window of an average weight (SW)
  Algorithm 2: Bipartition into halves and match syndromes (MB)
  Algorithm 3: Find and encode an error-free covering k-set (CS)

- **Results** for LDPC codes

  All three algorithms carry over to LDPC codes;
  All reduce DistVer complexity $2^{F(R)n}$ of linear codes
  Larger reductions hold for the Gallager's ensemble.

Exponents $F(R)$ for linear codes and LDPC (ℓ, m)-codes of rate $R = k/n = 1 - \ell/m$





Theorem. Let some ensemble of linear codes have length $n \to \infty$, distance δn, and relative erasure-correcting threshold $\rho = \rho(R)$. Then codewords of weight δn can be found with complexity exponents as follows:

Codes on GV bound $R = 1 - h(\delta_{GV})$	Any ensemble with δ and ρ
b: $F_{SW} = R(1-R)$	$F_{SW} = (1-\rho)h(\delta)$
c: $F_{MB} = (1-R)/2$	$F_{MB} = h(\delta)/2$
g: $F_{CS} = (1-R)(1-h[\delta_{GV}/(1-R)]$	$F_{CS} = h(\delta) - \rho h(\delta/\rho)$

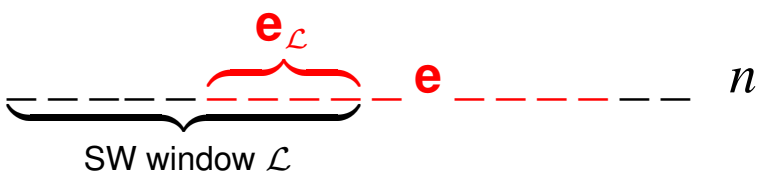
Sliding window (SW) technique for linear codes (Evseev 1983):

Any linear code C gives $P_{error}(C) < 2P_{ML}(C)$, by correcting δ_{GV} errors.

Note: Most LC have covering radius $\delta_{GV}(1+\varepsilon)$ (Blinovskii 1987).

Algorithm. Take any SW $\mathcal{L}$ of length $s \sim k + 2\log n$ in $[n,k]$ code.

Codeword **e** of weight d gives vector $\mathbf{e}_{\mathcal{L}}$ of weight $v \sim dR$ in some $\mathcal{L}$.



Take $d = 1, 2, \ldots$ Run $n\binom{s}{v}$ encoding trials for all $\mathbf{e}_{\mathcal{L}}$. STOP if encoded vector **e** has weight d. Then $F = R(1-R)$.

A. Algorithm works for all cyclic codes and most long linear codes.

B. For most linear codes, we can uniquely encode all n SWs on length $s = k + o(n)$. Equivalently, we correct $n - s$ erasures.

C. For LDPC codes, $s/k > 1$. We increase s to get unique encoding.

Matched Bipartition (MB) technique (Dumer, Stern; 1986* - 1989**)

* MB works for *any* linear code and has the lowest exponent as $R \to 1$.

** Combined with Covering Sets, MB reduces exponent F_{CS} for all R.

Algorithm. Take two disjoint $n/2$-windows $\mathcal{L}_1$ and $\mathcal{L}_2$. Some partition $\mathcal{L}_1, \mathcal{L}_2$ decouples any vector $\mathbf{e}$ of weight $d = 1, 2, ...$ into vectors $\mathbf{e_1}$ and $\mathbf{e_2}$ of wt $\sim d/2$.

Lists $\{\mathbf{e}_1\}$ and $\{\mathbf{e}_2\}$ have size $M \sim \binom{n/2}{d/2}$ for any linear $[n, k]$ code.

Calculate syndromes $h(\mathbf{e}_1)$ and $h(\mathbf{e}_2)$ for each $\mathbf{e}_1$ and $\mathbf{e}_2$.

Sort the list $\{\mathbf{e}_1\} \cup \{\mathbf{e}_2\}$ to find $\mathbf{e}_1, \mathbf{e}_2$ with $h(\mathbf{e}_1) = h(\mathbf{e}_2)$.

Output a codeword $\mathbf{c} = (\mathbf{e}_1, \mathbf{e}_2)$ if it exists and STOP.

Matching of $\{\mathbf{e}_1\}$ and $\{\mathbf{e}_2\}$ requires $\sim M \log_2 M \sim 2^{nh(\delta)/2} \sim 2^{n(1-R)/2}$ operations for classical codes on GV bound.

Covering Sets (Prange, Leon, Kruk, Coffey-Goodman..., 1962-1990)

A set $\mathcal{J}$ of θn positions is τ-deficient in code $C[n,k]$ if the generator submatrix $G_{[n]\setminus\mathcal{J}}$ has rank $k-\tau$. Then *shortened* code $C_{\mathcal{J}}$ has size 2^{τ} and erasure set $\mathcal{J}$ can be restored into some code list $M_{\mathcal{J}}$ of size $2^{\tau(\mathcal{J})}$.

Theorem: for most linear codes, all $(n-k)$-sets $\mathcal{J}$ have $\tau \leq \sqrt{2n}$.

ML decoding. Use $[n, n-k, d]$ covering of size $L \sim (n \ln n) \binom{n}{d} / \binom{n-k}{d}$.

Some $(n-k)$ set $\mathcal{J}$ covers error e of weight d with probability $1 - e^{-n \ln n}$.

Recover a code list $M_{\mathcal{J}}$ from erasure set $\mathcal{J}$ and find the closest codeword.

Let codes $C_{\mathcal{J}} \setminus 0$ of length θn have **average** size N_θ (over codes $C \in \mathbb{C}$ and sets $\mathcal{J}$). **The erasure threshold** is ρ if $N_\theta \to 0$ for $\theta < \rho$ and $N_\theta \geq 1$, $\theta > \rho$.

Lemma 1: Most codes $C \in \mathbb{C}$ correct most erasure sets $\mathcal{J}$ if $N_\theta \to 0$.

Lemma 2: $N_\theta = \sum_{\tau=0}^{\theta n} (2^\tau - 1)\, \alpha_\theta(\tau)$, where $\alpha_\theta(\tau)$ is the fraction of τ-def. θn-sets $\mathcal{J}$ in codes $C \in \mathbb{C}$. Most codes C have $\leq 2^{-\tau}$ fraction of τ-def. ρn-sets.

ML complexity (per trial). We need one Gaussian elimination and 2^τ vector add-s to recover τ-def. erasure set $\mathcal{J}$. This has complexity $\mathcal{D}_\theta(\mathcal{J}) \leq n^3 + n2^\tau$.

Recover θn-sets with ave. complexity $\mathcal{D}_\theta \leq \sum_{\tau=0}^{\theta n}(n^3 + n2^\tau)\alpha_\theta(\tau) = nN_\theta + n^3$.
If $N_\theta \to 0$, only $\leq 1/n$ of codes C have complexity $\mathcal{D} \geq n^4 L$ over L trials.

DistVer (per trial). To cover any codeword $c \neq 0$ of weight d with sets $\mathcal{J}$, we take $\mathcal{J}$ with $\tau(\mathcal{J}) \geq 1$. Again $\mathcal{D}_\theta \leq \sum_{\tau=1}^{\theta n}[n(2^\tau - 1) + n^3]\alpha_\theta(\tau) < nN_\theta + n^3$.

From generic to LDPC codes. We use two general decoupled procedures, ρn-erasure recovery and δn-covering sets. Parameters δ and ρ give the no. of trials L and complexity exponent $F_{CS} = (\log_2 L)\, n \sim h(\delta) - \rho h(\delta/\rho)$.

Parameters of two LDPC ensembles: Gallager 1963, Litsyn-Shevelev 2002

1. Ensemble $\mathbb{A}(\ell, m)$: all p.-check $r \times n$ matrices H with column weight ℓ and row weight $m = \ell n/r$. Code rate $R = 1 - \alpha$, where $\alpha = \ell/m$.

2. Ensemble $\mathbb{B}(\ell, m)$: H consists of ℓ horizontal blocks $H_1, ..., H_\ell$. Block H_1 includes m consecutive unit matrices of size $\frac{r}{\ell} \times \frac{r}{\ell}$. Any other block H_i is some random permutation $\pi_i(n)$ of n columns of H_1. Ensembles $\mathbb{A}(\ell, m)$ and $\mathbb{B}(\ell, m)$ have the best LDPC spectra known to date.

3. For any $\beta \in [0, 1]$, let $t > 0$ be the (single) root of the equation

$$\frac{(1+t)^{m-1} + (1-t)^{m-1}}{(1+t)^m + (1-t)^m} = 1 - \beta \text{ and } q(\beta) = \alpha \log_2 \frac{(1+t)^m + (1-t)^m}{2t^{\beta m}} - \alpha m h(\beta).$$

Lemma*: A vector of weight βn belongs to some code C of $\mathbb{A}, \mathbb{B}(\ell, m)$ with probability $\asymp 2^{-nq(\beta)}$. There are on average $N_\theta \asymp 2^{-nf(\theta)}$ nonzero vectors on any set J of size θn, where $f(\theta) = \max_{0<\beta\leq 1} \{q(\beta\theta) + \theta h(\beta)\}$.

Distance δ and threshold ρ are the roots of : $h(\delta) + q(\delta) = 0, \quad f(\rho) = 0.$

Summary for LDPC codes and improvements for the Gallager's ensemble

LDPC codes reduce both the distance δ and erasure threshold ρ of linear codes. The former factor prevails and reduces $F_{\mathsf{CS}} = h(\delta) - \rho h(\delta/\rho)$. This design holds for any (ir)regular LDPC or other ensemble with the known δ and ρ. However, we increase F_{CS} if we need to correct δ_{VG} errors in MLD.

F_{CS} can be reduced for the Gallager's ensemble $\mathbb{B}(\ell, m)$. Here the first* n/m parity checks have disjoint supports J_i of length m. They represent code $B(1, m)$ with **all-even** weight (AE) m-blocks on each J_i. We use AE m-blocks of total w-t θn to cover AE m-blocks of total w-t δn, and say that vectors of w-t θn form a **Code-covering** $\mathcal{B}(\theta, \delta)$ in $B(1, m)$.

1	**1**	**1**	**1**	**1**												
						1	**1**	**1**	**1**	**1**						
												1	**1**	**1**	**1**	**1**
–	–	–	–	–		–	–	–	–	–		–	–	–	–	–
⋆	⋆							⋆	⋆			⋆	⋆			err
●	●	●	●					●	●			●	●			cov

Let a codeword c leave s_i parity checks with $i = 0, 2, \ldots$ free positions.

There are only $N_S < (n/m)^m$ possible spectra $S = \{s_0, s_2, \ldots, s_m\}$.

However, covering size of $\mathcal{B}(\theta, \delta)$ depends on spectra S.

Example. Code $B(1, 8)$, $n = 16$, $\delta n = 8$, $\theta n = 12$.

$$H = \left| \begin{array}{cccccccccccccccc} \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & & & & & & & & \\ & & & & & & & & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} \end{array} \right|$$

A. $- \; - \; - \; - \quad \underbrace{- \; - \; - \; -}_{\text{4 free pos.}} \quad - \; - \; - \; - \quad \underbrace{- \; - \; - \; -}_{\text{4 free pos.}}$

B. $- \; - \; - \; - \quad - \; - \; - \; - \quad \underbrace{- \; - \; - \; - \; - \; - \; - \; -}_{\text{8 free pos.}}$

C. $- \; - \; - \; - \quad - \; - \quad \underbrace{- \; -}_{\text{2 free pos.}} \quad - \; - \quad \underbrace{- \; - \; - \; - \; - \; -}_{\text{6 free pos.}}$

A. $s_4 = 2, \; s_{\text{other}} = 0, \quad |\mathcal{B}(\theta, \delta)| = \binom{4}{2} \cdot \binom{4}{2} + 2 \cdot \binom{4}{4} \cdot \binom{4}{0} = 38$

B. $s_8 = 1, \; s_{\text{other}} = 0, \quad |\mathcal{B}(\theta, \delta)| = \binom{8}{4} = 70$

C. $s_2 = 1, \; s_6 = 1, \; s_{\text{other}} = 0, \; |\mathcal{B}(\theta, \delta)| = \binom{2}{2} \cdot \binom{6}{2} + \binom{2}{0} \cdot \binom{6}{4} = 30$

Given $S = \{s_0, s_2, ..., s_m\}$, cover c_S of w-t δn with AE vectors $b = (c_S, b')$ of w-t θn. Here AE vector b' have w-t $\theta n - \delta n$ on the rest $n - \delta n$ positions.

Let $N(\theta)$ and $N_S(\theta, \delta)$ be the number of AE vectors b and b'.

Theorem. Covering $\mathcal{B}(\theta, \delta)$ has expon. size $L_S(\theta, \delta) \precsim N(\theta) \ / \ \min_S N_S(\theta, \delta)$.

For $m \to \infty$, the number $L_S(\theta, \delta)$ has the same order as the order $L(\theta, \delta) \sim \binom{n}{\delta n} / \binom{n-\theta n}{\delta n}$ of generic (non-coding) covering $T(n, \theta n, \delta n)$.

For finite m, $L_S(\theta, \delta)$ reduces the order of $L(\theta, \delta)$.