Easy Problems I

Coding Theory Seminar

E. Martínez-Moro

imUVa Instituto de Investigación en Matemáticas

Universidad de Valladolid

References

Complexity issues in coding theory, in Handbook of Coding Theory (1998) by A. Barg.

Outline

Reference

LDPC codes

Iterated majority voting

Gallager's emsemble

Complexity

Regular Graphs

Regular-Graph codes

LDPC codes

Let $\mathcal{C}$ be a binary linear code and H its parity check matrix. Assume that each column of H has h ones and each row l ones, and $l > h$.

If $n \to \infty$ and $l, h \ll$ then in every check a vanishing part of coordinates is involved and every coordinate takes part in a vanishing part of checks. Thus they are called Low Density Parity Check Codes.

Gallager proved that a code from a random ensemble of LDPC codes comes close to the asymptotic G-V bound.

Thus we have nice codes, how to decode them?

Iterated majority voting

- Find a coordinate contained in more unsatisfied checks that satisfied ones. Invert it and recompute the syndrome.
- Repeat until no shuch coordinate is found or stop after $\mathcal{O}(\log n)$ rounds.

Iterated majority voting

- Find a coordinate contained in more unsatisfied checks that satisfied ones. Invert it and recompute the syndrome.
- Repeat until no shuch coordinate is found or stop after $\mathcal{O}(\log n)$ rounds.

Instituto de Matemáticas

Iterated majority voting

Let $\mathbf{e}$ be an error, $e = \mathrm{wt}(\mathbf{e})$ and $\mathbf{s} = H \cdot \mathbf{e}^T (= H \cdot \mathbf{e}^T)$.We assume that there are functions bounding the weight of the syndrome depending just on the number of errors

$$w_\star(e) \leq \mathrm{wt}(\mathbf{s}) \leq w^\star(e) = eh.$$

Note that the last equality comes from just taking the worst case case, i.e. the ones of each of the columns marked by a position of the error do not overlap the ones in the other columns marked by $\mathbf{e}$.

Iterated majority voting

Let $\mathbf{e}$ be an error, $e = \mathrm{wt}(\mathbf{e})$ and $\mathbf{s} = H\cdot\mathbf{e}^T (= H\cdot\mathbf{e}^T)$.We assume that there are functions bounding the weight of the syndrome depending just on the number of errors

$$w_\star(e) \leq \mathrm{wt}(\mathbf{s}) \leq w^\star(e) = eh.$$

Note that the last equality comes from just taking the worst case case, i.e. the ones of each of the columns marked by a position of the error do not overlap the ones in the other columns marked by $\mathbf{e}$.

Iterated majority voting

Let $\mathbf{e}$ be an error, $e = \mathrm{wt}(\mathbf{e})$ and $\mathbf{s} = H \cdot \mathbf{e}^T (= H \cdot \mathbf{e}^T)$.We assume that there are functions bounding the weight of the syndrome depending just on the number of errors

$$w_\star(e) \leq \mathrm{wt}(\mathbf{s}) \leq w^\star(e) = eh.$$

Note that the last equality comes from just taking the worst case case, i.e. the ones of each of the columns marked by a position of the error do not overlap the ones in the other columns marked by $\mathbf{e}$.

Iterated majority voting

Let $\mathbf{e}$ be an error, $e = \mathrm{wt}(\mathbf{e})$ and $\mathbf{s} = H \cdot \mathbf{e}^T (= H \cdot \mathbf{e}^T)$.We assume that there are functions bounding the weight of the syndrome depending just on the number of errors

$$w_\star(e) \leq \mathrm{wt}(\mathbf{s}) \leq w^\star(e) = eh.$$

Note that the last equality comes from just taking the worst case case, i.e. the ones of each of the columns marked by a position of the error do not overlap the ones in the other columns marked by $\mathbf{e}$.

Iterated majority voting

Lemma .- If $\mathrm{wt}(\mathbf{s}) > eh/2$ there exist a coordinate , i, such that

$$\mathrm{wt}(\mathbf{s} + \mathbf{h}_i) < \mathrm{wt}(\mathbf{s})$$

where $\mathbf{h}_i$ is the i-th column of the parity check matrix H.

In other words, flipping the i-th coordinate of the error $\mathbf{e}$ (or the i-th coordinate of the received vector $\mathbf{r}$ the weight of the syndrome decreases in one unit.)

Proof. Just take into account that the average number of unsatisfied checks per coordinate is $\mathrm{wt}(\mathbf{s})/e > h/2$. □

Iterated majority voting

Lemma .- If $\mathrm{wt}(\mathbf{s}) > eh/2$ there exist a coordinate , i, such that

$$\mathrm{wt}(\mathbf{s}+\mathbf{h}_i) < \mathrm{wt}(\mathbf{s})$$

where $\mathbf{h}_i$ is the i-th column of the parity check matrix H.

In other words, flipping the i-th coordinate of the error **e** (or the i-th coordinate of the received vector **r** the weight of the syndrome decreases in one unit.)

Proof. Just take into account that the average number of unsatisfied checks per coordinate is $\mathrm{wt}(\mathbf{s})/e > h/2$. □

Iterated majority voting

Lemma .- If $\mathrm{wt}(\mathbf{s}) > eh/2$ there exist a coordinate , i, such that

$$\mathrm{wt}(\mathbf{s} + \mathbf{h}_i) < \mathrm{wt}(\mathbf{s})$$

where $\mathbf{h}_i$ is the i-th column of the parity check matrix H.

In other words, flipping the i-th coordinate of the error $\mathbf{e}$ (or the i-th coordinate of the received vector $\mathbf{r}$ the weight of the syndrome decreases in one unit.)

Proof. Just take into account that the average number of unsatisfied checks per coordinate is $\mathrm{wt}(\mathbf{s})/e > h/2$. □

Iterated majority voting

Note that the asssumption of the previous lemma is satisfied if

$$w_\star(e) > \frac{eh}{2},$$

which is indeed true for the case $e = 1$ (Note that $e = 1$ implies $\mathbf{e} = \mathbf{e}_i$ the i-th coordinate vector and $\mathrm{wt}(\mathbf{s}) = \mathrm{wt}(\mathbf{h}_i) = h$).

Thus there exist a non-empty region $0 \leq e \leq e_0$ such that the inequality in red in this slide is satisfied for all the values of the weight e of the error.Let us define by e_0 the maximal number with such a property.

Iterated majority voting

Note that the asssumption of the previous lemma is satisfied if

$$w_\star(e) > \frac{eh}{2},$$

which is indeed true for the case $e = 1$ (Note that $e = 1$ implies $\mathbf{e} = \mathbf{e}_i$ the i-th coordinate vector and $\mathrm{wt}(\mathbf{s}) = \mathrm{wt}(\mathbf{h}_i) = h$).

Thus there exist a non-empty region $0 \leq e \leq e_0$ such that the inequality in red in this slide is satisfied for all the values of the weight e of the error. Let us define by e_0 the maximal number with such a property.

Iterated majority voting

Note that the asssumption of the previous lemma is satisfied if

$$w_\star(e) > \frac{eh}{2},$$

which is indeed true for the case $e = 1$ (Note that $e = 1$ implies $\mathbf{e} = \mathbf{e}_i$ the i-th coordinate vector and $\mathrm{wt}(\mathbf{s}) = \mathrm{wt}(\mathbf{h}_i) = h$).

Thus there exist a non-empty region $0 \leq e \leq e_0$ such that the inequality in red in this slide is satisfied for all the values of the weight e of the error.Let us define by e_0 the maximal number with such a property.

Iterated majority voting

For any vector of weight $e \leq e_0$ there exist a coordinate contained in more than $h/2$ unsatisfied checks such that if we invert it:

1. We reduce by one the weight of the syndrome,
2. we add or remove an error.

Thus, taking into account the picture in the whiteboard, in the worst case we go to the "south-east" neighbor. Thus for successful decoding, even in the worst case, we do not have to leave the region $\mathrm{wt}(\mathbf{e}) < e_0$, that is, the initial point must be below the line containing the point $P = (e_0, e_0 h/2)$ with slope -1.

Iterated majority voting

For any vector of weight $e \leq e_0$ there exist a coordinate contained in more than $h/2$ unsatisfied checks such that if we invert it:

1. We reduce by one the weight of the syndrome,
2. we add or remove an error.

Thus, taking into account the picture in the whiteboard, in the worst case we go to the "south-east" neighbor. Thus for successful decoding, even in the worst case, we do not have to leave the region $\mathrm{wt}(\mathbf{e}) < e_0$, that is, the initial point must be below the line containing the point $P = (e_0, e_0 h/2)$ with slope -1.

Iterated majority voting

For any vector of weight $e \leq e_0$ there exist a coordinate contained in more than $h/2$ unsatisfied checks such that if we invert it:

1. We reduce by one the weight of the syndrome,
2. we add or remove an error.

Thus, taking into account the picture in the whiteboard, in the worst case we go to the "south-east" neighbor. Thus for successful decoding, even in the worst case, we do not have to leave the region $\mathrm{wt}(\mathbf{e}) < e_0$, that is, the initial point must be below the line containing the point $P = (e_0, e_0 h/2)$ with slope -1.

Iterated majority voting

For any vector of weight $e \leq e_0$ there exist a coordinate contained in more than $h/2$ unsatisfied checks such that if we invert it:

1. We reduce by one the weight of the syndrome,
2. we add or remove an error.

Thus, taking into account the picture in the whiteboard, in the worst case we go to the "south-east" neighbor. Thus for successful decoding, even in the worst case, we do not have to leave the region $\mathrm{wt}(\mathbf{e}) < e_0$, that is, the initial point must be below the line containing the point $P = (e_0, e_0 h/2)$ with slope -1.

Let us denote e_1 the first coordinate of that line with $w^\star(e) = eh$, then e_1 is a lower bound on the number of correctable errors and we have that

Theorem .- For every vector **e** of weight

$$\mathrm{wt}(\mathbf{e}) < \left\lfloor \frac{e_0}{2} \frac{h+2}{h+1} \right\rfloor$$

the IMV algorithm performs a succesfully decoding.

Let us denote e_1 the first coordinate of that line with $w^\star(e) = eh$, then e_1 is a lower bound on the number of correctable errors and we have that

Theorem .- For every vector $\mathbf{e}$ of weight

$$\mathrm{wt}(\mathbf{e}) < \left\lfloor \frac{e_0}{2} \frac{h+2}{h+1} \right\rfloor$$

the IMV algorithm performs a succesfully decoding.

Gallager's emsemble

Take l copies of the identity matrix I_m and form a $m \times ml$ matrix. Permute randomly its columns. Repeat this independenly h times to form an $mh \times ml$ parity-check matrix of the LDPC code. As $m \to \infty$ this defines an ensemble of LDPC codes of growing length.

Theorem (Zyablov-Pinser).- For almost all codes in Gallager's ensemble the value e_0/n is > 0. If $R \to 1$, the fraction of erros corrected by IMV is nor less than $\delta_0(R)/22$.

$\delta_0(R)$ is the smallest root of $R = 1 - H_2(x)$ is the relative Gibert-Varshamov distance. Note that the GV distance is the maximum d_0 s.t.

$$|\mathcal{C}| \sum_{i=0}^{d_0-1} \binom{n}{i} \leq 2^n.$$

Gallager's emsemble

Take l copies of the identity matrix I_m and form a $m \times ml$ matrix. Permute randomly its columns. Repeat this independently h times to form an $mh \times ml$ parity-check matrix of the LDPC code. As $m \to \infty$ this defines an ensemble of LDPC codes of growing length.

Theorem (Zyablov-Pinser).- For almost all codes in Gallager's ensemble the value e_0/n is > 0. If $R \to 1$, the fraction of erros corrected by IMV is nor less than $\delta_0(R)/22$.

$\delta_0(R)$ is the smallest root of $R = 1 - H_2(x)$ is the relative Gibert-Varshamov distance. Note that the GV distance is the maximum d_0 s.t.

$$|\mathcal{C}| \sum_{i=0}^{d_0-1} \binom{n}{i} \leq 2^n.$$

Complexity

Theorem.- The IMV algorithm has complexity $\mathcal{O}(n \log n)$.

Proof. Assume that l, h are constants. Thus every chech involves l coordinates, i.e. $\mathcal{O}(1)$. One decoding round is performed in $\mathcal{O}(n)$ time thus one must show that each round reduces the syndrome weight by a finite fraction.

Let e be the number of errors, then c of them will be contained in more that $h/2$ unsatisfied checks. Therefore

$$s = \mathrm{wt}(\mathbf{s}) \leq ch + (e-c)\frac{h}{2}.$$

I.e.

$$c \geq \frac{2s - eh}{h}.$$

Complexity

Theorem.- The IMV algorithm has complexity $\mathcal{O}(n \log n)$.

Proof. Assume that l, h are constants. Thus every chech involves l coordinates, i.e. $\mathcal{O}(1)$. One decoding round is performed in $\mathcal{O}(n)$ time thus one must show that each round reduces the syndrome weight by a finite fraction.

Let e be the number of errors, then c of them will be contained in more that $h/2$ unsatisfied checks. Therefore

$$s = \mathrm{wt}(\mathbf{s}) \leq ch + (e - c)\frac{h}{2}.$$

I.e.

$$c \geq \frac{2s - eh}{h}.$$

Complexity

Theorem.- The IMV algorithm has complexity $\mathcal{O}(n \log n)$.

Proof. Assume that l, h are constants. Thus every chech involves l coordinates, i.e. $\mathcal{O}(1)$. One decoding round is performed in $\mathcal{O}(n)$ time thus one must show that each round reduces the syndrome weight by a finite fraction.

Let e be the number of errors, then c of them will be contained in more that $h/2$ unsatisfied checks. Therefore

$$s = \mathrm{wt}(\mathbf{s}) \leq ch + (e - c)\frac{h}{2}.$$

I.e.

$$c \geq \frac{2s - eh}{h}.$$

Complexity

Now a coordinate (say i_0) appears in h checks and each of this checks involves $l-1$ other coordinates. We shall thenote those coordinates connected to i_0.

Thus flipping a coordinate changes the value of h checks and affects at most to other $h(l-1)$ coordinates connected with it. Suppose that all of them where contained (before flipping) in more that $h/2$ unsatisfied checks and should have to be inverted in the present decoding round.

Complexity

Now a coordinate (say i_0) appears in h checks and each of this checks involves $l-1$ other coordinates. We shall thenote those coordinates connected to i_0.

Thus flipping a coordinate changes the value of h checks and affects at most to other $h(l-1)$ coordinates connected with it. Suppose that all of them where contained (before flipping) in more that $h/2$ unsatisfied checks and should have to be inverted in the present decoding round.

Complexity

Suppose also (we go for the worst case)that inverting the i_0-th coordinate change the status in such a way that flipping the connected ones does not reduce the weight of $\mathbf{s}$, thus the algorithm leaves them unchange.

Let us mark all the coordinates that need to be inverted during one decoding round, a single flip can remove at most

$$h(l-1)+1 \text{ marks.}$$

Complexity

Suppose also (we go for the worst case)that inverting the i_0-th coordinate change the status in such a way that flipping the connected ones does not reduce the weight of $\mathbf{s}$, thus the algorithm leaves them unchange.

Let us mark all the coordinates that need to be inverted during one decoding round, a single flip can remove at most

$$h(l-1)+1 \text{ marks.}$$

Complexity

Thus the number of coordinates inverted (during a round) is greater than

$$\frac{c}{h(l-1)+1}.$$

Inverting one coordinate reduces the weight of the syndrome in at least one unit, thus after a round

$$s-\frac{c}{h(l-1)+1}\leq s\left(1-\frac{2s-eh}{sh(h(l-1)+1)}\right)$$

and the expresion in red is <1 thus we are done. □

Complexity

Thus the number of coordinates inverted (during a round) is greater than

$$\frac{c}{h(l-1)+1}.$$

Inverting one coordinate reduces the weight of the syndrome in at least one unit, thus after a round

$$s-\frac{c}{h(l-1)+1}\leq s\left(1-\frac{2s-eh}{sh(h(l-1)+1)}\right)$$

and the expresion in red is < 1 thus we are done. □

Parallel iterated majority voting

- In parallel, mark all coordinates contained in more unsatisfied checks that satisfied ones. Invert them and recompute the syndrome.
- Repeat until no coordinates can be marked.

Parallel iterated majority voting

- In parallel, mark all coordinates contained in more unsatisfied checks that satisfied ones. Invert them and recompute the syndrome.
- Repeat until no coordinates can be marked.

Graphs

Let $G = (V_1 \cup V_2, E)$ be a bipartite graph. G is called (l, h)-regular if the degree of every vertex in V_1 is l and the degree of every vertex in V_2 is h, $h < l$.

Let $|V_2| = n$, thus $|V_1| = hn/l$. Now we form the matrix A with rows numbered by vertices from V_1 and column with vertices from V_2. Let $v_i \in V_1$ and $v'_j \in V_2$, then $a_{ij} = 1$ iff $(v_i, v'_j) \in E$.

Note that $n - k \leq nh/l$ thus $k > n - nh/l = n(1 - h/l)$ and thus A is the parity check matrix of a code $\mathcal{C}(G)$ of rate $R > 1 - h/l$.

Graphs

Let $G = (V_1 \cup V_2, E)$ be a bipartite graph. G is called (l, h)-regular if the degree of every vertex in V_1 is l and the degree of every vertex in V_2 is h, $h < l$.

Let $|V_2| = n$, thus $|V_1| = hn/l$. Now we form the matrix A with rows numbered by vertices from V_1 and column with vertices from V_2. Let $v_i \in V_1$ and $v_j' \in V_2$, then $a_{ij} = 1$ iff $(v_i, v_j') \in E$.

Note that $n - k \leq nh/l$ thus $k > n - nh/l = n(1 - h/l)$ and thus A is the parity check matrix of a code $\mathcal{C}(G)$ of rate $R > 1 - h/l$.

Graphs

Let $G = (V_1 \cup V_2, E)$ be a bipartite graph. G is called (l, h)-regular if the degree of every vertex in V_1 is l and the degree of every vertex in V_2 is h, $h < l$.

Let $|V_2| = n$, thus $|V_1| = hn/l$. Now we form the matrix A with rows numbered by vertices from V_1 and column with vertices from V_2. Let $v_i \in V_1$ and $v'_j \in V_2$, then $a_{ij} = 1$ iff $(v_i, v'_j) \in E$.

Note that $n - k \leq nh/l$ thus $k > n - nh/l = n(1 - h/l)$ and thus A is the parity check matrix of a code $\mathcal{C}(G)$ of rate $R > 1 - h/l$.

Graphs

Lemma .- Suppose that every group of $e \leq a = \alpha n$ h-regular vertices has at least $(3/4+\epsilon)he$ neighbors. Then IMV applied to $\mathcal{C}(G)$ corrects $\alpha n/2$ errors.

The key idea is that any not too large subset of h-regular vertices has not to few neighbors.

A bipartite graph is called a (l, h, α, γ)-expander if every subset $U \subseteq V_1$ of cardinal at most a fraction of h-regular vertices has at least $\gamma|U|$ neighbors,more formally

$$|U| < \alpha n \Rightarrow |\{u' \in V_2 \mid \exists u \in U\,,\,(u,u') \in E\}| \geq \gamma|U|.$$

Graphs

Lemma .- Suppose that every group of $e \leq a = \alpha n$ h-regular vertices has at least $(3/4+\epsilon)he$ neighbors. Then IMV applied to $\mathcal{C}(G)$ corrects $\alpha n/2$ errors.

The key idea is that any not too large subset of h-regular vertices has not to few neighbors.

A bipartite graph is called a (l, h, α, γ)-expander if every subset $U \subseteq V_1$ of cardinal at most a fraction of h-regular vertices has at least $\gamma|U|$ neighbors,more formally

$$|U| < \alpha n \Rightarrow |\{u' \in V_2 \mid \exists u \in U\,,\,(u,u') \in E\}| \geq \gamma|U|.$$

Instituto de Matemáticas

Graphs

Lemma .- Suppose that every group of $e \leq a = \alpha n$ h-regular vertices has at least $(3/4+\epsilon)he$ neighbors. Then IMV applied to $\mathcal{C}(G)$ corrects $\alpha n/2$ errors.

The key idea is that any not too large subset of h-regular vertices has not to few neighbors.

A bipartite graph is called a (l, h, α, γ)-expander if every subset $U \subseteq V_1$ of cardinal at most a fraction of h-regular vertices has at least $\gamma|U|$ neighbors,more formally

$$|U| < \alpha n \Rightarrow |\{u' \in V_2 \mid \exists u \in U\,,\,(u, u') \in E\}| \geq \gamma|U|.$$

Graphs

Proof .- We shall show that the expansion in the lemma ensures $e \leq e_0$.

The total number of checks with errors is (by asumption of the lemma)

$$\mathrm{wt}(\mathbf{s}) + x > \frac{3}{4}he$$

where x denotes the number of satisfied checks with errors.

Now, any unsatisfied check contains at least one error and every satisfied check with at least one error contains at least two errors, therefore

$$\mathrm{wt}(\mathbf{s}) + 2x \leq he.$$

Graphs

Proof .- We shall show that the expansion in the lemma ensures $e \leq e_0$.

The total number of checks with errors is (by asumption of the lemma)

$$\mathrm{wt}(\mathbf{s}) + x > \frac{3}{4}he$$

where x denotes the number of satisfied checks with errors.

Now, any unsatisfied check contains at least one error and every satisfied check with at least one error contains at least two errors, therefore

$$\mathrm{wt}(\mathbf{s}) + 2x \leq he.$$

Graphs

Now by the two expresions in red in the previous slide we get that $\mathrm{wt}(\mathbf{s}) > eh/2$ for any vector of weight up to $e \leq a$, thus $e < e_0$ by definition of the latter one. □

Note that the lemma corresponds to choosing the graph G so that $w_\star(e)$ is nearly $eh/2$ but unfortunatelly the level of expansion greater than $3/4$ is not really known for explicit constructions of families of expanders, though an average random bipartite graph could probably have a good expansion.

Graphs

Now by the two expresions in red in the previous slide we get that $\mathrm{wt}(\mathbf{s}) > eh/2$ for any vector of weight up to $e \leq a$, thus $e < e_0$ by definition of the latter one. □

Note that the lemma corresponds to choosing the graph G so that $w_\star(e)$ is nearly $eh/2$ but unfortunatelly the level of expansion greater than $3/4$ is not really known for explicit constructions of families of expanders, though an average random bipartite graph could probably have a good expansion.

Graphs

Theorem .- Let $G = (V_1 \cup V_2, E)$ a randomly chosen (l, h)-regular graph with $|V_2| = n$, $|V_1| = hn/l$. Then for all $0 < \alpha < 1$ a set of αn vertices in V_2 will have (on the average) at least

$$n\frac{h}{l}(1-(1-\alpha)^l)$$

neighbors.

Regular-Graph codes

An explicit construction:

Let $\mathcal{H}$ be a parity check matrix of a binary $[l, m, \beta l]$-code $\mathcal{C}$. A the $v \times n$ vertices vs. edges incidence matrix of a l regular graph. The parity check matrix H of the Regular-Graph code $\mathcal{C}(G, \mathcal{H})$ is obtained by replacing the ith row in A by its $l - m$ copies an then replacing the l all-one columns in these $l - m$ rows by l columns of $\mathcal{H}$, $1 \leq i \leq v$.

Note that $n - k \leq (l - m)v$, thus $1 - R \leq 2 - mv/n$ since $lv = 2n$, and therefore

$$mv/n - 1 = 2m/l - 1 \leq R.$$

See an example in the whiteboard.

Regular-Graph codes

An explicit construction:

Let $\mathcal{H}$ be a parity check matrix of a binary $[l, m, \beta l]$-code $\mathcal{C}$. A the $v \times n$ vertices vs. edges incidence matrix of a l regular graph. The parity check matrix H of the Regular-Graph code $\mathcal{C}(G, \mathcal{H})$ is obtained by replacing the ith row in A by its $l - m$ copies an then replacing the l all-one columns in these $l - m$ rows by l columns of $\mathcal{H}$, $1 \leq i \leq v$.

Note that $n - k \leq (l - m)v$, thus $1 - R \leq 2 - mv/n$ since $lv = 2n$, and therefore

$$mv/n - 1 = 2m/l - 1 \leq R.$$

See an example in the whiteboard.

Parallel Decoding R-G codes

- In parallel, for each of the v subsets, if the current setting of coordinates is within distance $\beta l/4$ of a codeword of $\mathcal{C}$, mark all the coordinated that should be flipped to get such a codeword. Invert the marked coordinates.
- Repeat $\mathcal{O}(\log n)$ rounds.