2. Γραμμικοί κώδικες

Ας υποθέσουμε τώρα ότι τα σύμβολα ελέγχου μπορούν να προκύψουν από το k-μήνυμα με τέτοιο τρόπο ώστε οι κωδικές λέξεις x να ικανοποιούν το σύστημα με γραμμικές εξισώσεις

$$Hx^T = 0,$$

όπου H είναι ένας δοσμένος $(n - k) \times n$ πίνακας με στοιχεία από το σώμα $\mathbf{F}_q$. Η **κανονική μορφή** για τον H είναι $[A \mid I_{n-k}]$ όπου A ένας $(n - k) \times k$ πίνακας και I_{n-k} ο $(n - k) \times (n - k)$ μοναδιαίος πίνακας.

Προκύπτει ο παρακάτω

Ορισμός 1.21 Έστω H ένας $(n - k) \times n$ πίνακας με βαθμό $n - k$ και στοιχεία από το σώμα $\mathbf{F}_q$. Το σύνολο όλων των n-διάστατων διανυσμάτων x που ικανοποιούν την εξίσωση $Hx^T = 0$ ονομάζονται **γραμμικός κώδικας** C πάνω από το $\mathbf{F}_q$ με **μήκος** n. Ο πίνακας H είναι ο **πίνακας ελέγχου ισοτιμίας (parity-check matrix)** του κώδικα C ο οποίος ονομάζεται και γραμμικός (n, k)-κώδικας. Αν ο H είναι στην μορφή $[A \mid I_{n-k}]$ τότε τα πρώτα k σύμβολα από την κωδική λέξη x είναι το αρχικό k-μήνυμα, ενώ τα υπόλοιπα $n - k$ σύμβολα του x είναι τα σύμβολα ελέγχου. Ο C ονομάζεται επίσης **συστηματικός γραμμικός (n, k)-κώδικας** και τότε θεωρούμε ότι ο H είναι στην **κανονική μορφή**. Αν $q = 2$ τότε ο C ονομάζεται **δυαδικός κώδικας** (**binary code**).

Παρατήρηση 2.1 Το σύνολο C των λύσεων x της $Hx^T = 0$ (ή αλλιώς ο **μηδενόχωρος** του H) είναι ένας υπόχωρος του διανυσματικού χώρου $\mathbf{F}_q^n$ με διάσταση k. Επειδή οι κωδικές λέξεις είναι προσθετική ομάδα, ο C ονομάζεται επίσης **κώδικας-ομάδα**.

Ιδιότητες γραμμικών κωδίκων 2.2

1. $x \in C \Leftrightarrow Hx^T = 0$
2. Ο πίνακας ελέγχου ισοτιμίας H του κώδικα C γράφεται σε κανονική μορφή $H = [A \mid I_{n-k}]$
3. Ο πίνακας $G = [I_k \mid -A^T]$ ονομάζεται (**κανονικός**) **γεννήτορας πίνακας** του κώδικα C με πίνακα ελέγχου ισοτιμίας $H = [A \mid I_{n-k}]$ στην κανονική μορφή λόγω του ότι ισχύει $x = uG$ οι κωδικές λέξεις είναι γραμμικοί συνδυασμοί των γραμμών του G.

Ισχύει: $GH^T = HG^T = 0$.

Παράδειγμα 2.3

$$\text{Αν } H = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \end{bmatrix} \text{ τότε } G = \begin{bmatrix} 1 & 0 & 0 & 1 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

Από τις τρεις γραμμές του πίνακα παίρνουμε τον (3, 6)-κώδικα $C \subseteq \mathbf{F}_2^6$ ο οποίος αποτελείται από 8 κωδικές λέξεις:

000000, 100110, 010101, 001011, 110011, 011110, 101101, 111000

Όπως και πριν κάθε κωδική λέξη x μπορεί να περιγραφεί σαν διανύσμα της μορφής $x = uG$, όπου $u = u_1u_2u_3$ με $u_i \in \mathbf{F}_2$.

Υπάρχουν τέσσερις κωδικές λέξεις βάρους 3, τρεις κωδικές λέξεις βάρους 4 και μια κωδική λέξη βάρους 0. Η ελάχιστη απόσταση του κώδικα είναι 3, επομένως ανιχνεύει την ύπαρξη το πολύ δύο λαθών και διορθώνει το πολύ ένα λάθος.

Θεώρημα 2.4 Έστω G ο γεννήτορας πίνακας ενός γραμμικού κώδικα C. Τότε οι γραμμές του G σχηματίζουν μια βάση του C.

Απόδειξη Οι k γραμμές του πίνακα G είναι γραμμικώς ανεξάρτητες από τον ορισμό του γεννήτορα πίνακα ενός γραμμικού κώδικα. Αν r είναι ένα διάνυσμα-γραμμή του G τότε $rH^T = 0$ άρα και $Hr^T = 0$ για κάθε $r \in C$. τώρα, dimC είναι η διάσταση του μηδενόχωρου του H, η οποία είναι $n - \text{rank}(H) = k$. Επομένως, οι k γραμμές του G σχηματίζουν μια βάση του C.

Ένας κώδικας μπορεί να έχει πολλούς πίνακες ελέγχου ισοτιμίας και γεννήτορες πίνακες. Κάθε $k \times n$ πίνακας του οποίου ο χώρος γραμμών είναι ίσος με τον C μπορεί να είναι επίσης ένας γεννήτορας πίνακας του C.

Αν ο «γεννήτορας πίνακας» H δεν είναι στην κανονική μορφή μπορούμε να τον μετατρέψουμε σε ένα πίνακα της μορφής $[I_k \mid -A^T]$ χωρίς να αλλάξουμε τον μηδενόχωρο του H, δηλαδή τον κώδικα C. Μετά μετατρέπουμε τις συντεταγμένες για να σχηματίσουμε τον πίνακα H' ο οποίος να είναι σε κανονική μορφή. Οι συντεταγμένες του κώδικα C' που αντιστοιχεί στον H' είναι «ισοδύναμος» με τον C με την ακόλουθη έννοια:

Ορισμός 2.5 Δύο κώδικες C και C' ίδιου μήκους n θα λέγονται **ισοδύναμοι** αν υπάρχει μια μετάθεση π του συνόλου $\{1, 2, \ldots, n\}$ τέτοια ώστε

$$(x_1, \ldots, x_n) \in C \Leftrightarrow \left(x_{\pi(1)}, \ldots, x_{\pi(n)}\right) \in C'$$

Έτσι, σχηματίζουμε τον γεννήτορα πίνακα G' του πίνακα C' και ύστερα εφαρμόζουμε την αντίστροφη μετάθεση π^{-1} στις συντεταγμένες.

Το επόμενο αποτέλεσμα μας δείχνει ότι για κάθε γραμμικό κώδικα, η ελάχιστη απόσταση μπορεί να υπολογισθεί από το βάρος Hamming των κωδικών λέξεων.

Μια από τις πιο σημαντικές ιδιότητες των γραμμικών κωδίκων είναι η παρακάτω

Πρόταση 2.6 Έστω C ένας γραμμικός (n, k)-κώδικας. Η ελάχιστη απόσταση του C είναι ίση με το ελάχιστο δυνατό βάρος που έχει κωδική λέξη διάφορη του μηδενικού στοιχείου.

Απόδειξη Έστω w το ελάχιστο δυνατό βάρος Hamming κωδικής λέξης διάφορης του μηδενικού στοιχείου 0. Έστω $x \in C$ μια κωδική λέξη βάρους Hamming w. Τότε ισχύει ότι $d(x, 0) = w(x) = w$. Επομένως, ισχύει ότι $w \geq d_{min}(C)$. Τώρα έστω u και v ένα ζευγάρι κωδικών λέξεων του C με απόσταση τέτοια ώστε $d(u, v) = d_{min}(C)$. Αφού C γραμμικός κώδικας έπεται ότι και η u – v είναι επίσης κωδική λέξη. Η u – v έχει βάρος $d_{min}(C)$. Επομένως, $d_{min}(C) \geq w$. Δηλαδή, $d_{min}(C) = w$.

Ορισμός 2.7 Ένα γραμμικός κώδικας C μήκους n, διάστασης k και ελάχιστης απόστασης d θα ονομάζεται **(n, k, d)-κώδικας**.

Έστω τώρα, $u = u_1, \ldots, u_n$ και $v = v_1, \ldots, v_n$ δύο διανύσματα του διανυσματικού χώρου $\mathbf{F}_q^n$ και έστω $u \cdot v = u_1v_1 + \ldots + u_nv_n$ να συμβολίζει το γινόμενο των u και v πάνω από τον $\mathbf{F}_q^n$. Αν $u \cdot v = 0$ τότε τα u και v θα λέγονται **ορθογώνια**.

Ορισμός 2.8 Έστω C ένας γραμμικός (n, k)-κώδικας ορισμένος στο σώμα $\mathbf{F}_q$. Ο **ορθογώνιος κώδικας** $C^{\perp}$ του κώδικα C ορίζεται να είναι ο

$$C^{\perp} = \{u \mid uv = 0 \text{ για κάθε } v \in C\}$$

Επειδή ο C είναι ένας k-διάστατος υπόχωρος του n-διάστατου διανυσματικού χώρου $\mathbf{F}_q^n$ το ορθογώνιο συμπλήρωμα του C είναι διάστασης n – k και είναι ένας (n, n – k) κώδικας. Μπορεί να αποδειχτεί ότι αν ο κώδικας C έχει γεννήτορα τον πίνακα G και πίνακα ελέγχου ισοτιμίας H τότε ο $C^{\perp}$ έχει γεννήτορα πίνακα τον H και πίνακα ελέγχου ισοτιμίας του G. Η ορθογωνιότητα των δύο κωδίκων μπορεί να εκφραστεί από τη σχέση $GH^T = HG^T = 0$. Τώρα θα συνοψίσουμε κάποιες απλές ιδιότητες των γραμμικών κωδίκων.

Παρατήρηση 2.9 Έστω mld(H) ο ελάχιστος αριθμός γραμμικά εξαρτημένων στηλών του H. Επειδή οποιεσδήποτε rank(H) + 1 το πλήθος στήλες του H είναι γραμμικά εξαρτημένες προφανώς ισχύει, $mld(H) \leq rank(H) + 1$ για κάθε πίνακα H.

Θεώρημα 2.10 Έστω H ένας πίνακας ελέγχου ισοδυναμίας ενός (n, k, d)-κώδικα C με $n > k$. Τότε ισχύουν:

(i) $\dim C = k = n - rank(H)$
(ii) $d = mld(H)$
(iii) $d \leq n - k + 1$.

Απόδειξη Το (i) είναι προφανές ενώ το (iii) προκύπτει από το (ii) και την προηγούμενη παρατήρηση. Για να αποδείξουμε το (ii) ας υποθέσουμε ότι ο H έχει στήλες $s_1, \ldots, s_n$. Παίρνουμε μια κωδική λέξη $c = (c_1, \ldots, c_n) \in C$ με βάρος w. Τότε επειδή

$$Hc^T = c_1s_1 + \ldots c_ns_n$$

ισχύει ότι $c_1s_1 + \ldots c_ns_n = 0$. Έχουμε επίσης ότι η c έχει μη-μηδενική συντεταγμένη σε w θέσεις επομένως κάποιες w, και μάλιστα όχι λιγότερες, στο πλήθος στήλες του H είναι γραμμικά εξαρτημένες. Δηλαδή, $mld(H) = w$. Εφαρμόζοντας την πρόταση 2.6

έχουμε ότι η ελάχιστη απόσταση d του κώδικα είναι ίση με το βάρος της c και συνεπώς το ζητούμενο.

Προκειμένου να επιβεβαιώσουμε την ύπαρξη γραμμικών (n, k)-κωδίκων με ελάχιστη απόσταση d πάνω από το $\mathbf{F}_q$ αρκεί να δείξουμε ότι υπάρχει $(n - k) \times n$ πίνακας H με $mld(H) = d$.

Αναφέρουμε δύο θεωρήματα χωρίς απόδειξη

Θεώρημα 2.11 (Φράγμα των Gilbert – Varshamov)

Αν

$$q^{n-k} > \sum_{i=0}^{d-2} \binom{n-1}{i} (q-1)^i$$

τότε μπορούμε να κατασκευάσουμε έναν γραμμικό (n, k)-κώδικα ορισμένο στο σώμα $\mathbf{F}_q$ με ελάχιστη απόσταση μεγαλύτερη ή ίση από d.

Θεώρημα 2.12 (Φράγμα του Plotkin) Αν υπάρχει ένας γραμμικός κώδικας μήκους n με M κωδικές λέξεις και ελάχιστη απόσταση d πάνω από το $\mathbf{F}_q$ τότε

$$d \leq n \frac{M(q-1)}{(M-1)q}$$