Do godziny 23:59 dnia **10 maja 2023 r.** wysyłają Państwo rozwiązanie projektu na maila **M.Zwierzynski@mini.pw.edu.pl** (w jednym pliku .PDF, każde zadanie na osobnej stronie/stronach, strony odpowiednio poobracane). Tytuł maila to **[AwAD 2022] Projekt nr 2**.

Kody liniowe

Wstęp teoretyczny

W poniższych zadaniach przyjmujemy, że wektory są w konwencji pionowej. Jeśli będziemy rozważać wektory w konwencji poziomej, to będą one określane przez transpozycję wektora zapisanego w konwencji pionowej.

(n,k)–**kodem liniowym** (nad ciałem $\mathbb{K}$) nazywamy podprzestrzeń liniową $\mathcal{C}$ wymiaru k przestrzeni liniowej $\mathbb{K}^n$ nad ciałem $\mathbb{K}$, gdzie $\mathbb{K}$ jest ciałem skończonym (czyli mającym skończoną liczność). Ciało $\mathbb{K}$ będziemy nazywać **alfabetem kodu**, zaś (n,k)–kod liniowy nad ciałem $\mathbb{K}$ będziemy określać **kodem** $|\mathbb{K}|$**-arnym**.

Dla (n,k)–kodu liniowego $\mathcal{C}$ dowolny wektor $v \in \mathcal{C}$ będziemy nazywać **słowem kodowym** kodu $\mathcal{C}$. Z kolei bazę przestrzeni $\mathcal{C}$ nazywać będziemy **bazą** kodu $\mathcal{C}$.

Dla dowolnego (n,k)–kodu liniowego $\mathcal{C}$ **macierzą generującą** kod $\mathcal{C}$ nazywamy macierz G o k wierszach i n kolumnach taką, że

$$G = \begin{pmatrix} e_1^T \\ e_2^T \\ \vdots \\ e_k^T \end{pmatrix},$$

gdzie $(e_1, e_2, ...e_k)$ jest bazą kodu $\mathcal{C}$. Oczywiście dla kodu liniowego $\mathcal{C}$ może istnieć więcej niż jedna macierz generująca, ponieważ może istnieć więcej niż jedna baza kodu $\mathcal{C}$.

Powyższe definicje zobrazujmy przykładem. Zbiór

$$\mathcal{C} = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 2 \\ 2 \end{pmatrix} \right\}.$$

jest podprzestrzenią przestrzeni liniowej $\mathbb{Z}_3^2$ nad skończonym ciałem $\mathbb{Z}_3$ wymiaru 1. A zatem $\mathcal{C}$ jest $(2,1)$–kodem liniowym nad ciałem $\mathbb{Z}_3$. Jest to kod 3-arny (terarny). Wektory przestrzeni $\mathcal{C}$ to słowa kodowe. Z kolei jedną z baz kodu $\mathcal{C}$ jest $B = \left((1,1)^T\right)$. A macierzą generującą powstałą z bazy B dla kodu $\mathcal{C}$ jest

$$G = \begin{pmatrix} 1 & 1 \end{pmatrix}.$$

Dla (n,k)-kodu liniowego $\mathcal{C}$ nad ciałem $\mathbb{K}$ i jego macierzy generującej G **kodowaniem** wektora $v \in \mathbb{K}^k$ nazywamy wektor $w \in \mathbb{K}^n$ taki, że

$$w = \left(v^T \cdot G\right)^T.$$

Wynikiem kodowania wektora v zawsze będzie słowo kodowe kodu $\mathcal{C}$.

Do zbadania jak różnią się różne wektory w teorii kodowania często używana jest odległość Hamminga. Niech V będzie dowolnym, niepustym podzbiorem przestrzeni liniowej $\mathbb{K}^n$ nad ciałem $\mathbb{K}$. **Odległością Hamminga** będziemy nazywać funkcję $d : V \times V \to \mathbb{R}$ daną wzorem

$$d(u,v) = |\{i \in [n] : u_i \neq v_i\}|$$

dla wektorów $u, v \in V$.

Oprócz procesu kodowania wektora może interesować nas proces dekodowania. Nie zawsze jednak wektor, który chcemy dekodować musi być słowem kodowym. W trakcie przesyłania zakodowanej wiadomości mogły przykładowo wystąpić zakłócenia transmisji. W związku z tym wynik dekodowania nie zawsze musi być wiadomością, którą zakodował nadawca. Jedną z wielu technik dekodowania wiadomości jest algorytm `MinimizeHammingDistance`. Załóżmy, że dysponujemy (n,k)–kodem liniowym $\mathcal{C}$ nad ciałem $\mathbb{K}$. Niech B będzie bazą kodu $\mathcal{C}$, z pomocą której utworzono macierz generującą G. Załóżmy, że mamy dany wektor $v \in \mathbb{K}^n$. Dekodowanie wektora v według tego algorytmu przebiega następująco:

`MinimizeHammingDistance(`$\mathcal{C}$`, `B`, `v`)`

IN: $\mathcal{C}$ – (n,k)–kod liniowy nad ciałem $\mathbb{K}$, B – baza kodu $\mathcal{C}$, v– dekodowany wektor

  $m = \min\{d(v,w) : w \in \mathcal{C}\}$ # d to odległość Hamminga

  $L = \{w \in \mathcal{C} : d(v,w) = m\}$

  w = losowo wybrany wektor należący do L

  r = wektor współczynników wektora w w bazie B

OUT: Wektor $r \in \mathbb{K}^k$

Jak już wspomnieliśmy – proces dekodowania nie zawsze musi dać nam w wyniku słowo, które było kodowane macierzą G na wejściu. W trakcie transmisji zakodowanego wektora mogły wystąpić błędy. Istnieją jednak oczywiście kody, które potrafią wykrywać błędy, a także je poprawiać. Na takich kodach nie będziemy jednak się skupiać w tym projekcie. Zainteresowani mogą sięgnąć do pozycji [1] z bibliografii.

Zadanka

Zadanka oznaczone symbolem $\star$ wymagają użycia wybranego języka programowania.

Zadanko 1. **Metryką** na zbiorze X nazywamy każdą funkcję $D : X^2 \to \mathbb{R}$ spełniającą następujące warunki:

- $D(x, y) = 0 \Leftrightarrow x = y$,
- $D(x, y) = D(y, x)$,
- $D(x, z) \le D(x, y) + D(y, z)$

dla $x, y, z \in X$. Udowodnić, że odległość Hamminga jest metryką.

Zadanko 2. Udowodnić, że dla dowolnego (n, k)–kodu liniowego $\mathcal{C}$ nad skończonym ciałem $\mathbb{K}$ i jego macierzy generującej G powstałej z bazy kodu B wynikiem kodowania dowolnego wektora $v \in \mathbb{K}^k$ jest słowo kodowe kodu $\mathcal{C}$.

Zadanko 3. Udowodnić, że dla dowolnego (n, k)–kodu liniowego $\mathcal{C}$ nad skończonym ciałem $\mathbb{K}$ i jego macierzy generującej G powstałej z bazy kodu B algorytm `MinimizeHammingDistance` użyty do dekodowania słowa kodowego $w \in \mathcal{C}$ zwróci taki wektor $v \in \mathbb{K}^k$, który w wyniku zakodowania go z użyciem macierzy G da wektor w.

Zadanko 4. Udowodnić, że dla dowolnej przestrzeni liniowej V nad ciałem $\mathbb{K}$ odległość Hamminga jest niezmiennicza ze względu na przesunięcia (czyli dla dowolnych wektorów $u, v, x \in \mathbb{K}^n$ odległość Hamminga słów u i v jest taka sama jak odległość słów $u + x$ i $v + x$).

Zadanko 5 ($\star$). Obliczyć odległość Hamminga dla wektorów $(1, 2, 0, 1)^T$ i $(0, 0, 0, 1)^T$. Które z wektorów ze zbioru

$$\left\{\begin{pmatrix}1\\2\\1\\2\\0\end{pmatrix}, \begin{pmatrix}1\\1\\1\\1\\1\end{pmatrix}, \begin{pmatrix}0\\0\\2\\1\\1\end{pmatrix}, \begin{pmatrix}2\\2\\2\\1\\0\end{pmatrix}\right\}$$

znajdują się najbliżej siebie w sensie Hamminga? W Mathematice przydać się może polecenie `HammingDistance[]`.

Zadanko 6 ($\star$). Wygenerować w wybranym języku wszystkie słowa kodowe dla $(5, 3)$-kodu liniowego $\mathcal{C}$ nad ciałem $\mathbb{Z}_7$ takiego, że bazą kodu liniowego $\mathcal{C}$ jest

$$B = \left(\begin{pmatrix}1\\0\\0\\2\\4\end{pmatrix}, \begin{pmatrix}0\\1\\0\\1\\0\end{pmatrix}, \begin{pmatrix}0\\0\\1\\5\\6\end{pmatrix}\right).$$

Zadanko 7 ($\star$). W wybranym języku utwórz macierz generującą G dla kodu liniowego $\mathcal{C}$ i bazy kodu B z zadanka 6. Następnie utwórz dowolnie wybrany przez siebie wektor $v \in \mathbb{Z}_7^5$ i wykonaj dekodowanie tego wektora używając algorytmu `MinimizeHammingDistance` dla kodu $\mathcal{C}$ i bazy B.

Zadanko 8 ($\star$). Celem tego zadania jest symulacja przesłania zakodowanej wiadomości. Do wykonania zadania użyj wybranego przez siebie języka.

a) Wygeneruj losową macierz o 10 kolumnach i 4 wierszach o wyrazach z ciała $\mathbb{Z}_5$.

b) Dokonaj unormowania macierzy z podpunktu a) do przedziału $[0, 1]$ dzieląc wszystkie wyrazy macierzy przez 4 (w tym podpunkcie potraktuj elementy macierzy jako liczby całkowite, a nie elementy z ciała $\mathbb{Z}_5$, czyli dzielenie przez 4 to standardowa operacja dzielenia na dwóch liczbach całkowitych). Na podstawie unormowanej macierzy utwórz obraz (przydatne w Mathematice może być polecenie `Image[]`)

c) Dana jest macierz

$$G = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 4 & 4 & 2 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 3 & 0 & 2 & 2 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 2 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 4 & 3 & 0 \end{pmatrix}$$

o wyrazach z ciała $\mathbb{Z}_5$. Udowodnij, że istnieje $(11, 4)$–kod liniowy $\mathcal{C}$ nad ciałem $\mathbb{Z}_5$ taki, że G jest macierzą generującą kodu $\mathcal{C}$.

d) Dla dowolnej kolumny v macierzy z podpunktu a) zakoduj wektor v używając macierzy generującej G.

e) Dla każdego zakodowanego wektora z podpunktu d) zasymuluj wysłanie go do pewnego użytkownika poprzez kanał, który dla przesyłanego wektora v dla każdej pozycji dodaje modulo 5 losową liczbę ze zbioru $\{0, 3\}$, przy czym prawdopodobieństwo dodania liczby 0 wynosi $0{,}95$, zaś prawdopodobieństwo dodania 3 jest równe $0{,}05$.

f) Dla każdego zakodowanego wektora po przesłaniu go przez kanał odkoduj ten wektor używając algorytmu `MinimizeHammingDistance`.

g) Z odkodowanych wektorów utwórz macierz odpowiadającą macierzy kodowanej z podpunktu a).

h) Porównaj macierze z podpunktu a) i g). Ile kolumn macierzy z podpunktu a) zostało poprawnie odkodowanych?

i) Wyrazy odkodowanej macierzy z podpunktu g) unormuj do przedziału $[0, 1]$ analogiczną metodą jak w podpunkcie b). Następnie dla unormowanej macierzy utwórz obraz analogicznie jak w podpunkcie b).

Bibliografia

[1] Władysław Mochnacki, **Kody korekcyjne i kryptografia**, Oficyna Wydawnicza Politechniki Wrocławskiej, 2000