4 Perfektní a MDS kódy

Ať $x \in \mathbb{A}^n$, kde $|\mathbb{A}| = q$. Pro každé $r \geq 1$ je $S(x,r) = \{y \in \mathbb{A}^n;\, d(x,y) \leq r\}$ okolí bodu x, které se skládá ze všech vektorů, jež se od x liší v nanejvýš r pozicích. Je-li C kód, který opravuje až r-násobné chyby, tak pro u, $v \in C$, $u \neq v$ platí $S(u,r) \cap S(v,r) = 0$. Označme $V_q(n,r)$ velikost $S(u,r)$.

Lemma 4.1. *Ať C je $(n,h,d)_q$ kód a ať $2r < d$. Potom*

- $h \cdot V_q(n,r) \leq q^n$
- $V_q(n,r) = \sum_{i=0}^{r} (q-1)^r \binom{n}{i}$.

Důkaz. Množiny $S(x,r)$, $x \in C$, jsou po dvou disjunktní a mají velikost $V_q(n,r)$. Zjevně $V_q(n,r) - V_q(n,r-1) = (q-1)^r \binom{n}{r}$ pro každé $r \geq 1$, neboť pro změnu na r místech lze vybrat $\binom{n}{r}$ podmnožin a v každé měněné pozici je k dispozici $q-1$ možností. □

Z lemmatu plyne tzv. **Hammingova nerovnost**

$$|C| \cdot V_q\left(n, \left\lfloor \frac{d-1}{2} \right\rfloor\right) \leq q^n,$$

která pro $[n,k,d]_q$ kód má tvar

$$V_q\left(n, \left\lfloor \frac{d-1}{2} \right\rfloor\right) \leq q^{n-k}.$$

Ať C je q-ární kód délky n nad abecedou $\mathbb{A}$. Kód C se nazývá **r-perfektní**, jestliže $\bigcup_{x \in C} S(x,r) = \mathbb{A}^n$ a $S(x,r) \cap S(y,r) = 0$, pro všechna x, $y \in C$, $x \neq y$. Má-li kód C alespoň dva prvky, tak je zřejmé, že je r-perfektní právě když Hammingova nerovnost je rovností. V takovém případě zjevně $d = 2r+1$.

Kódu se říká **perfektní**, je-li r-perfektní pro nějaké (jednoznačně určené) r. Totální kód $\mathbb{A}^n$ je 0-perfektní, jednobodové kódy jsou n-perfektní. Perfektní je také opakovací binární kód liché délky $2r+1$. O těchto kódech mluvíme jako o triviálních perfektních kódech.

Hammingův kód má parametry $[n, n-\ell, 3]_q$, kde $n = \frac{q^\ell - 1}{q-1}$ a $\ell \geq 2$. Máme $V_q(n,1) = 1 + (q-1)n = q^\ell$ a $\ell = n - (n-\ell)$, takže vidíme, že Hammingovy kódy jsou také perfektní.

Jedním z největších úspěchů matematické teorie samoopravných kódů je důkaz, že jediné netriviální r-pefektní kódy pro $r \geq 2$ mají parametry $[23, 12, 7]_2$ a $[11, 6, 5]_3$. Jsou známy jako (perfektní) **binární Golayův kód** a (perfektní) **ternární Golayův kód**.

Zde se omezíme na důkaz existence a jednoznačnosti binárního Golayova kódu. Pro každý $(n,h,d)_q$ kód C nad $\mathbb{F}_q$ definujme jeho **váhový polynom** f_C jako $\sum f_i x^i \in \mathbb{Z}[x]$, kde f_i udává počet kódových slov váhy i. Vidíme, že f je nejvýše

stupně n a že $f_i \leq (q-1)^i \binom{n}{i}$. Je rovněž patrné, že $f_0 \in \{0,1\}$, přičemž $f_0 = 1$ právě když C obsahuje nulový vektor. Je-li C lineární kód váhy d, tak $f_0 = 1$ a $f_1 = f_2 = \ldots = f_{d-1} = 0$. Tohoto faktu lze využít pro výpočet váhových polynomů Golayových kódů.

Všimněme si, že $V_2(23,3) = 1+23+\binom{23}{2}+\binom{23}{3} = 1+23(1+11+77) = 2048 = 2^{23-12}$. Každý $[23,12,7]$ kód dává v Hammingové nerovnosti rovnost, a tudíž musí být perfektní.

Lemma 4.2. *Váhový polynom* $[23,12,7]$ *kódu je roven* $1 + 253x^7 + 506x^8 + 1288x^{11} + 1288x^{12} + 506x^{15} + 253x^{16} + x^{23}$.

Důkaz. Víme, že $f_0 = 1$ a $f_1 = f_2 = \cdots = f_6 = 0$. Uvažme $u \in \mathbb{F}_2^{23}$, $|u| = 4$. Vektor u leží v $S(c,3)$ pro nějaké kódové slovo c. Toto slovo je určené jednoznačně a má váhu nejvýše 7. Z $f_4 = f_5 = f_6 = 0$ plyne $|c| = 7$. Kódových slov váhy 7 je f_7 a každé slovo váhy 4 je ve vzdálenosti 3 od některého z nich. Proto platí $\binom{7}{3} f_7 = \binom{23}{4}$, odkud $f_7 = 253$. Úvahou o rozmístění slov délky 5 dostaneme $\binom{7}{2} f_7 + \binom{8}{3} f_8 = \binom{23}{5}$, odkud $f_8 = 506$. Při výpočtu f_9 je třeba být trochu opatrnější, protože slova váhy 6 lze ze slov váhy 7 obdržet nejen odstraněním jednoho bitu, ale také odstraněním dvou bitů a přidáním třetího na jiné pozici. Máme

$$\left(7 + \binom{7}{2}(23-7)\right) f_7 + \binom{8}{2} f_8 + \binom{9}{3} f_0 = \binom{23}{6}.$$

Snadno odsud vypočítáme, že $f_9 = 0$. Obdobným postupem, který již zde podrobně provádět nebudeme, se spočítají hodnoty f_i pro zbylé indexy i, kde $10 \leq i \leq 23$. □

Je-li C binární $[n,k,d]$ kód, tak jeho **rozšířením** rozumíme $[n+1,k]$ kód, kde v přidané pozici (bývá to ta poslední, někdy ale též první) je součet bitů z původních pozic. Rozšířený kód leží v paritním kódu. Pokud již C ležel v paritním kódu, tak v přidané pozici jsou pouze nuly. Proto se při rozšiřování předpokládá, že C obsahuje kódová slova liché délky.

Lemma 4.3. *Ať* C *je* $[n,k,d]$ *kód s paritní maticí* H*. Rozšířený kód* C' *má pro* d *liché minimální váhu* $d+1$*, zatímco pro* d *sudé se minimální váha nemění. Paritní matici* H' *kódu* C' *lze získat přidáním k* H *nulového sloupce a poté řádku složeného ze samých jedniček.*

Důkaz. Pro $u \in C$ ať $u' \in C'$ vznikne z u přidáním paritního bitu. Pak $|u'| = |u| + 1$ právě když $|u|$ je liché. Jinak $|u'| = |u|$. Zbytek je snadný. □

Je-li $C \subseteq \mathbb{F}_q^n$ perfektní kód, tak $u + C$ je také perfektní kód pro každé $u \in \mathbb{F}_q^n$. Proto při zkoumání perfektních kódů $C \subseteq \mathbb{F}_q^n$ lze předpokládat, že $0 \in C$.

Věta 4.4. *Až na permutační ekvivalenci existuje jediný* $(23, 4096, 7)$ *kód a jediný* $(24, 4096, 8)$ *kód, které obsahují nulový vektor. Oba jsou lineární. První z nich je perfektní a druhý je samoduální a má váhový polynom* $1 + 759x^8 + 2456x^{12} + 759x^{16} + x^{24}$.

Důkaz. Uvažme nějaký $(24, 2048, 8)$ kód C, kde $0 \in C$. Propíchnutím vznikne $(23, 2048)$ kód, který má váhu 8 nebo 7. Z Hammingova odhadu plyne, že propíchnutý kód musí být perfektní a že má váhu 7. Váhový kód propíchnutého kódu je popsán v Lemmatu 4.2.

Kdyby v kódu C bylo slovo liché váhy, tak vhodným propíchnutím dostaneme kód s jiným váhovým polynomem, a proto takové slovo existovat nemůže. Všechna slova v propíchnutém kódu tedy mají původ ve slově sudé délky, a proto polynom $1 + 759x^8 + 2456x^{12} + 759x^{16} + x^{24}$ skutečně je váhovým polynomem kódu C.

Každý kód $u+C$, kde $u \in C$, má nutně stejný váhový polynom, neboť je rovněž $(24, 4096, 8)$ kódem, který obsahuje 0. Podle Důsledku 2.14 je C samoduálním lineárním dvojnásobně sudým kódem. Víme, že C obsahuje slova váhy 12 i 24, a proto je podle Tvrzení 3.14 určen až na permutační ekvivalenci jednoznačně. Podle téhož tvrzení je až na permutační ekvivalenci určen jednoznačně i každý jeho propíchnutý kód. Takovým kódem přitom musí být každý $(24, 4096, 7)$ kód obsahující 0. Jeho váhový kód totiž známe z Lemmatu 4.2, takže víme, že jeho rozšířením vždy získáme $(24, 4096, 8)$ kód (operaci rozšíření lze zjevně použít i pro nelineární kódy). □

Perfektní kódy jsou největší možné z hlediska Hammingova odhadu. V kapitole 2 jsme definovali MDS kódy. Ty jsou největší možné lineární kódy z hlediska Singletonova odhadu $d \leq n - k + 1$. Podle důsledku 2.8 je $[n, k, d]_q$ kód MDS právě když každých $n - k$ sloupců libovolné prověrkové matice je lineárně nezávislých.

Tvrzení 4.5. *Každý $[n, k, d]_q$ kód C je MDS právě když $C^\perp$ je MDS.*

Důkaz. Ať G je generující matice kódu C. Pak $d < n - k + 1$ právě když G lze zvolit tak, aby v ní existoval řádek, který má k nul. Protože $k = n - (n - k)$ a protože G je podle Důsledku 2.8 prověrkovou maticí kódu $C^\perp$, tak ten nemůže být MDS kódem.

Pokud naopak $C^\perp$ není MDS, tak $C^\perp$ má prověrkovou matici, v níž lze nalézt k lineárně závislých sloupců. Jinými slovy, C má generující matici s k lineárně závislými sloupci. Úpravami na řádcích lze získat jinou generující matici s řádkem, ve kterém oněch k sloupců vytýká samé nuly, odkud $d < n - k + 1$. □

Kód s generující maticí $(1 \cdots 1)$ je zjevně vždy MDS. Jeho doplňkem je paritní kód. MDS kódy libovolné délky vždy existují. Nyní nahlédněme, že toto již neplatí, budeme-li předpokládat, že $3 \leq d \leq n - 1$.

Tvrzení 4.6. *Ať $[n, k, d]_q$ kód C je MDS a ať $3 \leq d \leq n - 1$. Potom je $d \leq q$ a $n - q + 1 \leq k \leq q - 1$.*

Důkaz. Ať kód C má předpokládané vlastnosti. Z $d = n - k + 1$ plyne $k + 1 = n + 2 - d$, takže podle Tvrzení 4.5 stejné vlastnosti má i kód duální.

A $G = (I\ A)$ je generující matice kódu C. Pokud G obsahuje nulu na pozici (i, j) uvnitř A (tedy $j > k$), tak vezmeme v úvahu prvých k sloupců vyjma i-tého a přidáme k nim j-tý sloupec. Tyto sloupce jsou lineárně závislé, což by byl spor. Matice A obsahuje tedy samé nenuly a má $n-k = d-1 \geq 2$ sloupců. Ať jsou $(a_1, \ldots, a_k)^\mathsf{T}$ a $(b_1, \ldots, b_k)^\mathsf{T}$ po řadě prvý a druhý sloupec matice A. Je-li $k \geq q$, tak jistě existují $1 \leq i < j \leq k$, že $a_i/b_i = a_j/b_j$. Pokud těmito sloupci nahradíme i-tý a j-tý sloupec matice I, dostaneme k sloupců, které jsou lineárně závislé. To nelze, a proto $k < q$, $n - k < q$ a $d = n - k + 1 \leq q$. □

Lemma 4.7. *Ať C je $[n, k]_q$ kód a ať G je jeho generující matice. Pak C je MDS právě když každých k sloupců G je lineárně nezávislých. Pokud C' obdržíme z C propíchnutím v nejvýše $n - k$ sloupcích, bude C' opět MDS kód.*

Důkaz. Prvá část plyne z Tvrzení 4.5 a Důsledku 2.8. Po propíchnutí v nejvýše $n - k$ sloupcích získáme matici, rozměrů $k \times n'$, kde $n' \geq k$ a kde je každých k sloupců lineárně nezávislých. □