Uvedení do lineárních kódů

Odpřednesenou látku naleznete v dodatku I skript *Abstraktní a konkrétní lineární algebra*.

Dnešní přednáška

1. Základní geometrické myšlenky teorie lineárních kódů.
2. Generující a kontrolní matice lineárního podprostoru.

Dobré zdroje dalších informací

1. Richard Wesley Hamming (1915–1998): Bellovy laboratoře, ∼1946, technika pro opravu chyb na děrných štítcích
2. J. Adámek, *Foundations of coding*, John Wiley & Sons, New York, 1991
3. D. J. C. MacKay, *Information theory, inference and learning algorithms*, Cambridge Univ. Press, 2003
4. W. C. Huffman a V. Pless, *Fundamentals of error-correcting codes*, Cambridge Univ. Press, 2003

Příští přednáška

1. Základy ortogonální geometrie v prostorech tvaru $\mathbb{F}^n$ nad $\mathbb{F}$, kde $\mathbb{F}$ je obecné těleso.

Kódování versus šifrování

1. Kódování: dvě strany (Alice a Bob) si vyměňují zprávy. Při přenosu zpráv může dojít k poškození vyslané zprávy.

   Předpokládejme, že Alice píše Bobovi. Chceme umožnit Bobovi opravit poškozenou zprávu bez nutnosti zpětného dotazu Alice.

   Můžeme použít metody lineární algebry: lineární kódy.

2. Šifrování: dvě strany (Alice a Bob) si vyměňují zprávy. Při přenosu zpráv nemůže dojít k poškození vyslané zprávy, ale může dojít k odposlechu třetí stranou (ta se jmenuje Eve[a]).

   Předpokládejme, že Alice píše Bobovi. Chceme takovou komunikaci, kterou Eve nedokáže efektivně přečíst.

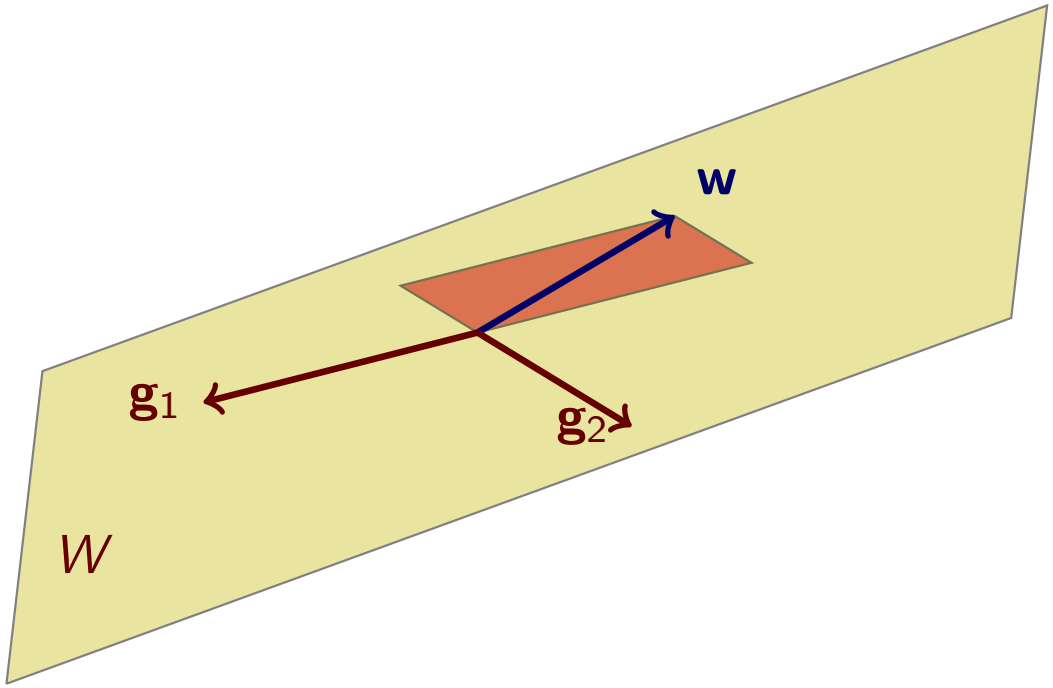
   K účinnému šifrování je třeba použít sofistikovaných metod. Viz např. J. Velebil, *Diskrétní matematika*, Praha, 2007.

[a] Z anglického *eavesdropper* — ten, kdo tajně naslouchá.

Rovina v $\mathbb{R}^3$ jako lineární kód

Rovina $x + y - z = 0$ je lineární podprostor W dimense 2 v $\mathbb{R}^3$.

1. Volbou uspořádané báze W lze generovat prvky W.
   1. W má usp. bázi (např.) $(\mathbf{g}_1, \mathbf{g}_2)$, kde $\mathbf{g}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}$, $\mathbf{g}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$.
   2. Tudíž $\mathbf{w} \in W$ iff existují jednoznačně určená $a_1, a_2 \in \mathbb{R}$ tak, že $a_1 \cdot \mathbf{g}_1 + a_2 \cdot \mathbf{g}_2 = \mathbf{w}$. (Protože báze určuje systém souřadnic.)



Rovina v $\mathbb{R}^3$ jako lineární kód (pokrač.)

Rovina $x + y - z = 0$ je lineární podprostor W dimense 2 v $\mathbb{R}^3$.

3. Vektor $\begin{pmatrix} a_1 \\ a_2 \end{pmatrix}$ budeme považovat za vektor informačních bitů.

   Neboli: volbou a_1, a_2 lze vygenerovat $\mathbf{w} \in W$ takto:

$$\underbrace{\begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}}_{\text{generující matice } \mathbf{G}} \cdot \begin{pmatrix} a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} a_1 \\ 2a_1 + a_2 \\ 3a_1 + a_2 \end{pmatrix} = \mathbf{w}$$

   Vektor $\mathbf{w}$ Alice odešle Bobovi.

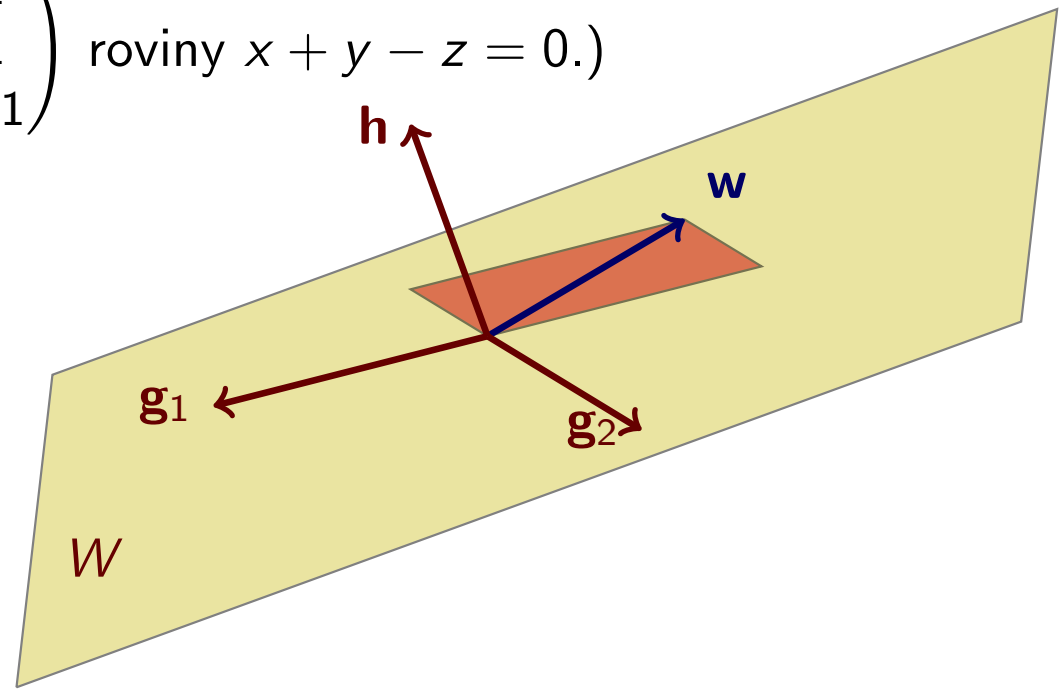
   Vektor $\mathbf{w}$ obsahuje redundantní informaci.[a] Tato redundantní informace chrání původní informační bity před poškozením.

[a] Podíl délky informace a celkové délky kódového slova (tzv information rate kódu) je tedy v našem případě $\frac{2}{3}$.

Rovina v $\mathbb{R}^3$ jako lineární kód (pokrač.)

Rovina $x + y - z = 0$ je lineární podprostor W dimense 2 v $\mathbb{R}^3$.

2. Volbou ortogonálního doplňku W lze testovat, zda vektory leží ve W.[a]
   1. W má ortogonální doplněk. Tudíž $\mathbf{w} \in W$ iff $\mathbf{h}^T \cdot \mathbf{w} = \mathbf{o}$. (Protože ortogonální doplněk je dán normálovým vektorem $\mathbf{h} = \begin{pmatrix} 1 \\ 1 \\ -1 \end{pmatrix}$ roviny $x + y - z = 0$.)



[a] Co je ortogonální doplněk vysvětlíme přesně v příští přednášce. Zatím se odvoláváme na intuici v $\mathbb{R}^3$.

Rovina v $\mathbb{R}^3$ jako lineární kód (pokrač.)

2. Neboli: syndrom s vektoru $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$, kde

$$s = \underbrace{\begin{pmatrix} 1 & 1 & -1 \end{pmatrix}}_{\text{kontrolní matice}} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$$

určuje míru příslušnosti vektoru $\mathbf{v}$ do W.

Povšimněme si: syndrom s vektoru $\mathbf{v}$ je hodnota standardního skalárního součinu $\langle \mathbf{h} \mid \mathbf{v} \rangle = \mathbf{h}^T \cdot \mathbf{v}$ v $\mathbb{R}^3$.

Syndrom vektoru $\mathbf{v}$ je tedy nulový právě tehdy, když jsou vektory $\mathbf{v}$ a $\mathbf{h}$ ortogonální.[a]

[a] V příští přednášce zobecníme pojem ortogonality vzhledem ke standardnímu skalárnímu součinu z prostorů $\mathbb{R}^n$ na prostory tvaru $\mathbb{F}^n$, kde $\mathbb{F}$ je obecné těleso.

Rovina v $\mathbb{R}^3$ jako lineární kód (pokrač.)

Rovina $x + y - z = 0$ je lineární podprostor W dimense 2 v $\mathbb{R}^3$.

Generující a kontrolní matice: $\mathbf{G} = \begin{pmatrix} 1 & 0 \\ 2 & 1 \\ 3 & 1 \end{pmatrix}$, $\mathbf{h}^T = \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}$.

Alice z informace $\begin{pmatrix} 3 \\ 2 \end{pmatrix}$ vygeneruje kódové slovo $\mathbf{G} \cdot \begin{pmatrix} 3 \\ 2 \end{pmatrix} = \begin{pmatrix} 3 \\ 8 \\ 11 \end{pmatrix}$ z prostoru W. Toto slovo odešle Bobovi.

Bob přijme slovo $\begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix}$. Došlo k poškození? Bob spočte syndrom přijatého slova:

$$\mathbf{h}^T \cdot \begin{pmatrix} 3 \\ 7 \\ 11 \end{pmatrix} = -1$$

Syndrom je nenulový, k chybě došlo. Na jaké posici k chybě došlo? Jak ji opravit?

Nearest neighbour decoding

Pokud jsme nepřijali kódové slovo, chceme najít kódové slovo, které je (v nějakém smyslu) nejblíže[a] přijatému slovu. Přijaté slovo pak nahradíme tímto nejbližším kódovým slovem.

[a] Zatím jde jen o slogan; v příští přednášce zavedeme Hammingovu vzdálenost (kódových) slov.

Problémy při opravě v lineárních kódech nad $\mathbb{R}$

1. Základní problém při nearest neighbour decoding nad $\mathbb{R}$: reálných čísel je příliš mnoho.
2. Potřebujeme „konečné číselné obory“, které se chovají stejně jako $\mathbb{R}$. Neboli: potřebujeme obecná konečná tělesa.[a]

   Důvod: chceme použít lineární algebru.

[a] Potřebujeme dostatečnou zásobu konečných těles $\mathbb{F}$. Existence nekonečně mnoha konečných těles souvisí s existencí nekonečného počtu prvočísel — viz příští přednášku.

Příklad: kód 10-ISBN

Deset cifer: použity jsou symboly z množiny $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, X\}$. Chápeme je jako zbytky po dělení číslem 11.

Příklad:

$$0-141-01878-X$$

kde jednotlivé skupiny znamenají:

1. 0 jazyk knihy (angličtina)
2. 141 nakladatelství (Penguin Mathematics)
3. 01878 číslo knihy, přidělené nakladatelstvím
4. X kontrolní bit

Obecně: kódové slovo kódu 10-ISBN je $x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}$, kde $\sum_{i=1}^{10} ix_i = 0$ jako zbytek po dělení číslem 11.

Kód 10-ISBN (pokrač.)

Kdy je řetězec $x_1x_2x_3x_4x_5x_6x_7x_8x_9x_{10}$ kódem ISBN?
Právě tehdy, když jeho syndrom

$$\underbrace{(1,2,3,4,5,6,7,8,9,X)}_{\text{kontrolní matice } \mathbf{H}^T \text{ kódu 10-ISBN}} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix}$$

je nula (počítáno jako zbytek po dělení číslem 11).[a]

[a] Povšimněme si: ISBN chápeme jako vektor v $(\mathbb{Z}_{11})^{10}$. Píšeme je tedy do sloupce.

Kód 10-ISBN (pokrač.)

Jak vytvořit kód ISBN?

Info o knize[a] = 9 bitů. Jak spočítat kontrolní bit?

$$\underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{pmatrix}}_{\text{generující matice } \mathbf{G} \text{ kódu 10-ISBN}} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \\ x_{10} \end{pmatrix}$$

počítáno jako zbytek po dělení číslem 11.

[a] Povšimněme si: informační bity chápeme jako vektor v $(\mathbb{Z}_{11})^9$. Píšeme je tedy do sloupce.

Kód 10-ISBN (pokrač.)

1. Kódy 10-ISBN = vektory v lineárním podprostoru W lineárního prostoru $(\mathbb{Z}_{11})^{10}$.
   Uspořádaná báze B prostoru W = sloupce matice $\mathbf{G}$.
   Dimense $W = 9$.
2. Info o knize = vektor souřadnic $\mathbf{a}$ vektoru $\mathbf{w}$ ve W vzhledem k uspořádané bázi B. Platí vztah $\mathbf{w} = \mathbf{G} \cdot \mathbf{a}$.
3. Test při příjmu slova $\mathbf{v}$ = výpočet syndromu $\mathbf{H}^T \cdot \mathbf{v}$ slova $\mathbf{v}$.
   Sloupce $\mathbf{H}$ = báze ortogonálního doplňku k W.

Kód 10-ISBN = lineární 11-kód délky 10 a dimense 9.

Kód 10-ISBN je schopen detekovat jednu chybu a prohození dvou pozic,[a] viz Příklad 3.3.2 textu *Diskrétní matematika*.

[a] To jsou běžné písařské chyby. 10-ISBN je starý kód, začíná být nahrazován kódem 13-ISBN.

Lineární podprostory prostoru $\mathbb{F}^n$ nad $\mathbb{F}$ (znovu a mírně jinak)

Ať W je lineární podprostor prostoru $\mathbb{F}^n$ nad $\mathbb{F}$. Víme, že platí $0 \leq \dim(W) \leq n$. Předpokládejme, že $\dim(W) = k > 0$.

1. Zvolme uspořádanou bázi $(\mathbf{g}_1, \ldots, \mathbf{g}_k)$ prostoru W. Označme jako $\mathbf{G} : \mathbb{F}^k \to \mathbb{F}^n$ matici se sloupcovým zápisem $(\mathbf{g}_1, \ldots, \mathbf{g}_k)$.

   Podle věty o dimensi jádra a obrazu platí $\mathrm{rank}(\mathbf{G}) = k$ a $\mathrm{def}(\mathbf{G}) = 0$. Navíc platí $\mathrm{im}(\mathbf{G}) = W$.

   Z toho okamžitě plyne:

   1. Pro jakoukoli volbu vektoru $\mathbf{a}$ z $\mathbb{F}^k$ je $\mathbf{G} \cdot \mathbf{a}$ vektor ve W.
   2. Vektor $\mathbf{w}$ z $\mathbb{F}^n$ leží ve W právě tehdy, když soustava rovnic $(\mathbf{G} \mid \mathbf{w})$ má právě jedno řešení (označme je $\mathbf{a}$).

      Toto jediné řešení $\mathbf{a}$ z $\mathbb{F}^k$ je vektor souřadnic vektoru $\mathbf{w}$ vzhledem k uspořádané bázi $(\mathbf{g}_1, \ldots, \mathbf{g}_k)$.

   Matici $\mathbf{G}$ říkáme generující matice[a] lineárního podprostoru W.

[a] Generující matice podprostoru W není jednoznačně určena: volbou jiné báze podprostoru W získáme jinou generující matici.

Lineární podprostory prostoru $\mathbb{F}^n$ nad $\mathbb{F}$ (pokrač.)

Ať W je lineární podprostor prostoru $\mathbb{F}^n$ nad $\mathbb{F}$ s uspořádanou bází $(\mathbf{g}_1, \ldots, \mathbf{g}_k)$, $k > 0$.

2. Protože $W = \mathrm{span}(\mathbf{g}_1, \ldots, \mathbf{g}_k)$, existuje soustava rovnic tvaru $(\mathbf{H}^T \mid \mathbf{o})$ tak, že řešení $(\mathbf{H}^T \mid \mathbf{o})$ je přesně W.[a]

   Podle Frobeniovy věty platí $\mathrm{rank}(\mathbf{H}^T) = n - k$ a $\mathrm{def}(\mathbf{H}^T) = k$.
   Víme, že rozměry $\mathbf{H}^T$ můžeme volit tak, aby platilo $\mathbf{H}^T : \mathbb{F}^n \to \mathbb{F}^{n-k}$.

Matici $\mathbf{H}^T$ říkáme kontrolní matice[b] lineárního podprostoru W.

Důvod: vektor $\mathbf{w}$ leží ve W právě tehdy, když $\mathbf{H}^T \cdot \mathbf{w} = \mathbf{o}$. Maticí $\mathbf{H}^T$ tedy kontrolujeme přítomnost vektoru v podprostoru W.

---

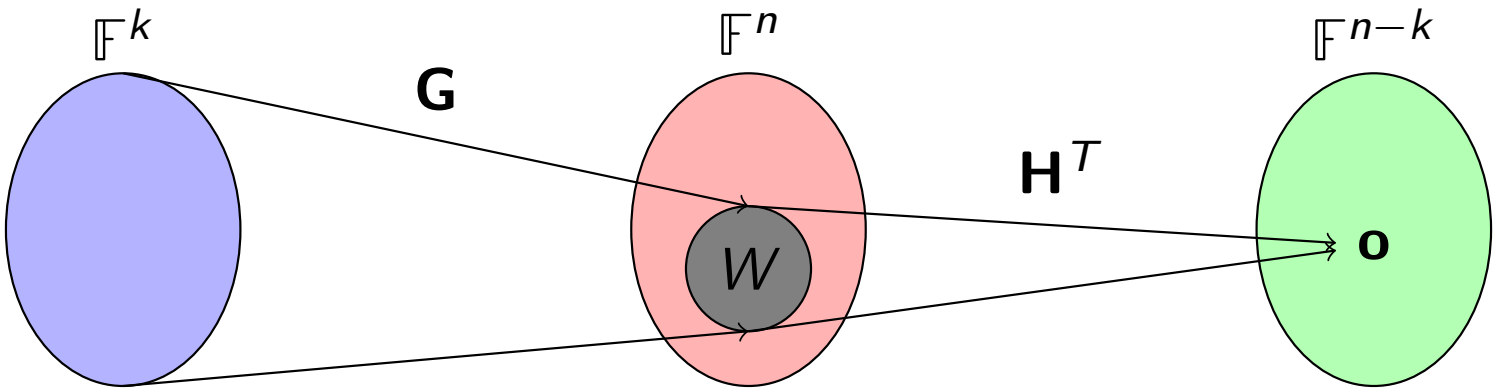
[a] Slogan: $\mathbf{H}$ je normála podprostoru W. To je důvod, proč píšeme v soustavě $(\mathbf{H}^T \mid \mathbf{o})$ matici soustavy jako transponovanou.

[b] Kontrolní matice podprostoru W není jednoznačně určena: volbou jiné soustavy rovnic získáme jinou kontrolní matici.

Základní vztah matic $\mathbf{G}$ a $\mathbf{H}^T$ pro lineární podprostor W

Platí

$$\begin{aligned}\{\mathbf{w} \in \mathbb{F}^n \mid \mathbf{G} \cdot \mathbf{a} = \mathbf{w} \text{ pro něj. } \mathbf{a} \text{ v } \mathbb{F}^k\} &= \mathrm{im}(\mathbf{G}) \\ &= W \\ &= \ker(\mathbf{H}^T) \\ &= \{\mathbf{w} \in \mathbb{F}^n \mid \mathbf{H}^T \cdot \mathbf{w} = \mathbf{o} \text{ v } \mathbb{F}^{n-k}\}\end{aligned}$$



Jinými slovy: $\mathbf{H}^T \cdot \mathbf{G} = \mathbf{O}_{k,n-k}$ a $\mathrm{rank}(\mathbf{G}) = \mathrm{def}(\mathbf{H}^T)$.

Co bude následovat v další přednášce?

1. V prostorech tvaru $\mathbb{F}^n$ zavedeme vztah ortogonality. To nám umožní mluvit přesně o generujících a kontrolních maticích lineárních podprostorů prostoru $\mathbb{F}^n$.
2. V prostorech tvaru $\mathbb{F}^n$ zavedeme pojem Hammingovy vzdálenosti vektorů. To nám později umožní zformulovat přesně metody detekce a opravy chyb v lineárních kódech.
3. Ukážeme, že existuje nekonečně mnoho konečných těles tvaru $\mathbb{Z}_p$, kde p je prvočíslo.

A co v následujících přednáškách?

Nahlédneme do teorie lineárních kódů.

1. Prostudujeme Hammingův $(7, 4)$-kód.
2. Ukážeme, jak Hammingova vzdálenost souvisí s detekcí a opravou chyb.
3. Ukážeme, jak vytvořit kontrolní matici z generující matice (a naopak).