Stavové prostory a minimalita

Bez ohledu na svoji fyzickou realizaci je kódování definováno svým chováním, což vede k pojmu *abstraktního stavu kódovače*. Abstraktní stav s lze charakterizovat jako zobrazení $\varphi_s : \mathbb{F}[[D]]^b \to \mathbb{F}[[D]]^c$ popisující, jak se kódovač v daném stavu bude chovat na příslušných vstupech. Je zřejmé, že různému chování bude při libovolné realizaci kódovače odpovídat jiný fyzický stav, zatímco naopak to při nevhodné realizaci platit nemusí.

Pro $\mathbf{u} \in \mathbb{F}((D))$ označme $\mathcal{K}(\mathbf{u}) := \sum_{i\geq 0} u_i D^i$ nezápornou (*kauzální*) část $\mathbf{u}$, a $\mathcal{Z}(\mathbf{u}) := \sum_{i<0} u_i D^i$ část zápornou (*antikauzální*). Toto značení přirozeně (po složkách) rozšíříme i na prvky $\mathbb{F}((D))^n$. Zřejmě platí $\mathbf{u} = \mathcal{Z}(\mathbf{u}) + \mathcal{K}(\mathbf{u})$. Jak $\mathcal{K}$ tak $\mathcal{Z}$ jsou přitom lineární zobrazení nad $\mathbb{F}$, ačkoli ne nad $\mathbb{F}(D)$ nebo $\mathbb{F}((D))$.

Počáteční stav kódovače se vyznačuje tím, že příslušné zobrazení φ_0 zobrazuje nulu na nulu a $\varphi_0(\mathbf{u})$ je prvkem kódu pro libovolné $\mathbf{u}$. V úvahu přicházejí ty stavy, do kterých se kódovač může dostat nějakým (předchozím) vstupem. Jinou charakterizací abstraktního stavu je tedy antikauzální prvek $\mathbf{w} = \mathcal{Z}(\mathbf{w})$. Zobrazení $\varphi_{\mathbf{w}}$ odpovídající takovému prvku je definováno vztahem

$$\varphi_{\mathbf{w}}(\mathbf{u}) = \mathcal{K}(\mathbf{w}\mathbf{G} + \mathbf{u}\mathbf{G}) = \mathcal{K}(\mathbf{w}\mathbf{G}) + \mathcal{K}(\mathbf{u}\mathbf{G}) = \mathcal{K}(\mathbf{w}\mathbf{G}) + \varphi_0(\mathbf{u})\,.$$

Z toho je vidět, že stav kódovače s daným nulovým stavem je jednoznačně určen hodnotou $\mathcal{K}(\mathbf{w}\mathbf{G}) = \varphi_{\mathbf{w}}(\mathbf{0})$.

Pro danou matici $\mathbf{G}$ definujme zobrazení $\mathcal{S}_{\mathbf{G}} : \mathbb{F}((D))^b \to \mathbb{F}[[D]]^c$

$$\mathcal{S}_{\mathbf{G}}(\mathbf{u}) := \mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G})\,.$$

To je lineární zobrazení vektorových prostorů nad $\mathbb{F}$ (ačkoli opět ne nad $\mathbb{F}(D)$ nebo $\mathbb{F}((D))$). Jádro tohoto zobrazení, tedy taková $\mathbf{u}$, pro která je $\mathcal{Z}(\mathbf{u})\mathbf{G}$ antikauzální, označme $\mathcal{S}^*_{\mathbf{G}}$.

Množina stavů kódovače definovaného maticí $\mathbf{G}$ (nebo krátce množina stavů matice $\mathbf{G}$) tedy odpovídá faktorprostoru

$$\Sigma_{\mathbf{G}} := \mathbb{F}((D))^b / \mathcal{S}^*_{\mathbf{G}}.$$

Řečeno slovy: stav kódovače v čase nula je dán vstupní posloupností, přičemž $\mathbf{u}$ a $\mathbf{u}'$ reprezentují stejný stav pokud jejich rozdíl $\mathbf{u} - \mathbf{u}'$ definuje nulový stav, tedy leží v jádru $\mathcal{S}$, tedy $\mathcal{K}(\mathcal{Z}(\mathbf{u} - \mathbf{u}')\mathbf{G}) = \mathbf{0}$.

Poznamenejme, že ve smyslu věty o isomorfismu můžeme za abstraktní stavy $\mathbf{G}$ považovat obrazy zobrazení $\mathcal{S}_{\mathbf{G}}$, tedy řady $\mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G})$. Formálně jsou ovšem prvky $\Sigma_{\mathbf{G}}$ rozkladové třídy $[\mathbf{u}]_{\mathbf{G}} := \mathbf{u} + \mathcal{S}^*_{\mathbf{G}}$.

Jak už jsme si všimli, různým abstraktním stavům musí odpovídat různé stavy kódovače. Můžeme tedy definovat minimální kódovač, jehož množina stavů je $\Sigma_{\mathbf{G}}$ a přechodová funkce je

$$([\mathbf{u}]_{\mathbf{G}}, \overrightarrow{u}) \mapsto ([D^{-1}(\mathcal{Z}(\mathbf{u}) + \overrightarrow{u})]_{\mathbf{G}}, \overrightarrow{v}),$$

kde $\overrightarrow{v}$ je absolutní člen Laurentovy řady $(\mathcal{Z}(\mathbf{u}) + \vec{u})\mathbf{G}$. (Tato konstrukce je podobná konstrukci minimálního deterministického automatu přijímajícího daný jazyk.)

*

Podobně jako je abstraktní stav kódovače definován jeho chováním pomocí vstupní posloupnosti, lze i bez znalosti kódovacího zobrazení definovat *(abstraktní) stav kódu*, jakožto množiny posloupností výstupních. Je-li $\mathbf{v} \in \mathcal{C}$, chápeme abstraktní stav kódu v čase nula odpovídající $\mathbf{v}$ jako stav (neznámého) kódovače, který právě

vyslal posloupnost $\mathcal{Z}(\mathbf{v})$ na základě nějaké vstupní posloupnosti w. Jedno z možných pokračování tohoto výstupu je $\mathcal{K}(\mathbf{v})$. Také v tomto případě bychom chtěli nějak popsat další možné výstupy. Bez ohledu na fyzický stav kódovače a nyní i bez ohledu na vstup víme, že možné budoucí výstupy kódovače jsou právě všechny posloupnosti $\mathbf{v}' \in \mathbb{F}[[D]]$, pro které je $\mathcal{Z}(\mathbf{v}) + \mathbf{v}' \in \mathcal{C}$. Je zde opět paralela se stavy minimálního konečného automatu. Pokud pro dvě kódová slova $\mathbf{v}_1$ a $\mathbf{v}_2$ existuje slovo $\mathbf{v}' \in \mathbb{F}[[D]]$ takové, že $\mathcal{Z}(\mathbf{v}_1) + \mathbf{v}' \in \mathcal{C}$, zatímco $\mathcal{Z}(\mathbf{v}_2) + \mathbf{v}' \notin \mathcal{C}$, pak libovolný automat s výstupem $\mathcal{C}$ musí být po vyslání slova $\mathcal{Z}(\mathbf{v}_1)$ v odlišném stavu než po vyslání slova $\mathcal{Z}(\mathbf{v}_2)$.

Definujme tedy na $\mathcal{C}$ zobrazení $\mathcal{S}_\mathcal{C}$ předpisem

$$\mathcal{S}_\mathcal{C} : \ \mathbf{v} \mapsto \{\mathbf{v}' \in \mathcal{C} \mid \mathcal{Z}(\mathbf{v}) + \mathcal{K}(\mathbf{v}') \in \mathcal{C}\}\,.$$

Označme

$$\mathcal{C}^* = \mathcal{S}_\mathcal{C}(\mathbf{0}) = \{\mathbf{w} \in \mathcal{C} \mid \mathcal{K}(\mathbf{w}) \in \mathcal{C}\},$$

což je zjevně podprostor $\mathcal{C}$ nad $\mathbb{F}$. Dostáváme následující fakt.

Tvrzení. $\mathcal{S}_\mathcal{C}$ je přirozená projekce $\mathcal{C}$ na $\mathcal{C}/\mathcal{C}^*$.

Důkaz. Pokud mají množiny $\mathcal{S}_\mathcal{C}(\mathbf{v}_1)$ a $\mathcal{S}_\mathcal{C}(\mathbf{v}_2)$ neprázdný průnik, pak $\mathcal{Z}(\mathbf{v}_1 - \mathbf{v}_2) \in \mathcal{C}$, a tedy také $\mathcal{K}(\mathbf{v}_1 - \mathbf{v}_2) \in \mathcal{C}$, neboli $\mathbf{v}_1 - \mathbf{v}_2 \in \mathcal{C}^*$. Pokud naopak $\mathcal{Z}(\mathbf{v}_1) - \mathcal{Z}(\mathbf{v}_2) \in \mathcal{C}$, pak pro každé $\mathbf{v}'$ platí, že $\mathcal{Z}(\mathbf{v}_1) + \mathcal{K}(\mathbf{v}') \in \mathcal{C}$, právě když $\mathcal{Z}(\mathbf{v}_2) + \mathcal{K}(\mathbf{v}') \in \mathcal{C}$. Navíc zřejmě $\mathbf{v} \in \mathcal{S}_\mathcal{C}(\mathbf{v})$ pro každé $\mathbf{v} \in \mathcal{C}$. □

Množinu

$$\Sigma_\mathcal{C} := \mathcal{C}/\mathcal{C}^*$$

s prvky $[\mathbf{v}]_\mathcal{C} = \mathbf{v} + \mathcal{C}^*$ nazýváme množinou *abstraktních stavů kódu*.

Vztah mezi abstraktními stavy matice $\mathbf{G}$ a abstraktními stavy jejího kódu je přirozeně dán lineárním zobrazením:

$$\begin{aligned} \mathcal{G} : \Sigma_\mathbf{G} &\to \Sigma_\mathcal{C} \\ [\mathbf{u}] &\mapsto [\mathbf{u}\mathbf{G}]\,. \end{aligned}$$

Ověřme, že toto zobrazení je definováno korektně, tj. že z $[\mathbf{u}]_\mathbf{G} = [\mathbf{u}']_\mathbf{G}$ plyne $[\mathbf{u}\mathbf{G}]_\mathcal{C} = [\mathbf{u}'\mathbf{G}]_\mathcal{C}$. Nechť tedy $\mathbf{w} = \mathbf{u} - \mathbf{u}' \in \mathcal{S}^*_\mathbf{G}$. Pak

$$\mathcal{K}(\mathbf{w}\mathbf{G}) = \mathcal{K}(\mathcal{Z}(\mathbf{w})\mathbf{G}) + \mathcal{K}(\mathcal{K}(\mathbf{w})\mathbf{G})) = \mathbf{0} + \mathcal{K}(\mathbf{w})\mathbf{G} \in \mathcal{C},$$

a tedy $\mathbf{u}\mathbf{G} - \mathbf{u}'\mathbf{G} \in \mathcal{C}^*$, jak jsme chtěli. (Použili jsme, že $\mathcal{K}(\mathcal{K}(\mathbf{w})\mathbf{G})$ je díky realizovatelnosti matice $\mathbf{G}$ rovno $\mathcal{K}(\mathbf{w})\mathbf{G}$.)

Opět platí, že pro různé množiny možných výstupů musí být kódovač v různých stavech. Tentokrát to ovšem platí pro jakýkoli (tedy i minimální) kódovač *jakéhokoli* zobrazení. Odtud plyne, že $\dim \Sigma_\mathcal{C} \leq \dim \Sigma_\mathbf{G}$, přičemž rovnost nastává právě když je zobrazení $\mathcal{G}$ prosté. V takovém případě řekneme, že je matice $\mathbf{G}$ *minimální*. Při hledání minimální matice jsme v situaci, kdy máme kód, tedy obraz kódování, a hledáme takové kódování, tedy takové přiřazení vzorů obrazům, které bude umožňovat nejmenší možný kódovač, jaký je pro daný kód vůbec možný.

V tuto chvíli ještě není zřejmé, že kódování, pro které je $\mathcal{G}$ prosté, vždy existuje, ale dále uvidíme, že tomu tak je. Následující věta podává různé charakteristiky minimálního kódování.

Věta. *Nechť je $\mathbf{G}$ generující matice. Následující podmínky jsou ekvivalentní:*

(1) $\mathbf{G}$ je minimální.

(2) $\mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) \in \mathcal{C}$, právě když $\mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) = \mathbf{0}$.

(3) Pro každé $\mathbf{u}$ *platí*
   a) $\deg \mathbf{u} \le \deg \mathbf{uG}$,
   b) $\operatorname{del} \mathbf{u} = \operatorname{del} \mathbf{uG}$.

(4) $\mathbf{G}$ *má polynomiální pravý inverz a současně pravý inverz polynomiální v* D^{-1}.

Podmínka (2) bývá formulována jako „pouze nulový abstraktní stav je kódovým slovem“. Všimněme si zejména, že $\mathbf{0} \in \mathcal{C}$ vždy, takže podmínku lze redukovat na přímou implikaci. Podmínka (3) říká, že rozsah obrazu je větší než rozsah vzoru. Minimalita je tedy porušena, právě když je možné aby zpráva začínala později nebo končila dříve než vstup.

Důkaz. „(1) $\Leftrightarrow$ (2)“: Zobrazení $\mathcal{G}$ je prosté, právě když platí, že

$$\mathcal{K}(\mathbf{uG}) \in \mathcal{C} \quad \text{právě když} \quad \mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) = \mathbf{0}.$$

Zbývá si všimnout, že $\mathcal{K}(\mathbf{uG}) \in \mathcal{C}$ je ekvivalentní $\mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) \in \mathcal{C}$. To plyne z toho, že

$$(*) \qquad \mathcal{K}(\mathbf{uG}) = \mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) + \mathcal{K}(\mathbf{u})\mathbf{G}\,.$$

„(2) $\Leftarrow$ (3)“: Předpokládejme, že (2) neplatí, tedy že pro nějaké $\mathbf{w} = \mathcal{Z}(\mathbf{w})$ máme $\mathbf{0} \neq \mathcal{K}(\mathbf{wG}) \in \mathcal{C}$ a nechť $\mathcal{K}(\mathbf{wG}) = \mathbf{uG}$. Tedy $0 \le \operatorname{del} \mathbf{uG}$, a je-li $\operatorname{del} \mathbf{u} < 0$, je porušeno (3)b). Pokud $0 \le \operatorname{del} \mathbf{u}$, platí $0 \le \deg \mathbf{u} = \deg(\mathbf{u} - \mathbf{w})$, a protože $(\mathbf{u}-\mathbf{w})\mathbf{G} = \mathcal{Z}(\mathbf{wG})$, dostáváme $\deg(\mathbf{u}-\mathbf{w})\mathbf{G} < 0 \le \deg(\mathbf{u}-\mathbf{w})$, čímž je porušeno (3)a) pro $\mathbf{u} - \mathbf{w}$.

„(2) $\Rightarrow$ (3)“:

a) Nechť pro nějaké $\mathbf{u}$ platí $\deg \mathbf{uG} < \deg \mathbf{u}$. Můžeme bez újmy na obecnosti předpokládat $\deg \mathbf{uG} = -1$, tj. $\mathcal{K}(\mathbf{uG}) = \mathbf{0}$. Pak je $\mathcal{K}(\mathbf{u})$, a tedy i $\mathcal{K}(\mathbf{u})\mathbf{G}$ nenulové a s pomocí $(*)$ máme

$$\mathbf{0} \neq -\mathcal{K}(\mathbf{u})\mathbf{G} = \mathcal{K}(\mathbf{uG}) - \mathcal{K}(\mathbf{u})\mathbf{G} = \mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) \in \mathcal{C}\,.$$

b) Nechť pro nějaké $\mathbf{u}$ platí $\operatorname{del} \mathbf{u} < \operatorname{del} \mathbf{uG}$. Bez újmy na obecnosti předpokládejme, že $\operatorname{del} \mathbf{uG} = 0$. Pak, opět s pomocí $(*)$,

$$\mathbf{0} \neq \mathcal{Z}(\mathbf{u})\mathbf{G} = \mathbf{uG} - \mathcal{K}(\mathbf{u})\mathbf{G} = \mathcal{K}(\mathbf{uG}) - \mathcal{K}(\mathbf{u})\mathbf{G} = \mathcal{K}(\mathcal{Z}(\mathbf{u})\mathbf{G}) \in \mathcal{C}\,.$$

„(3) $\Rightarrow$ (4)“: Použijeme charakterizaci existence polynomiálního inverzu.

Je-li $\mathbf{uG}$ polynomiální, je díky b) polynomiální i $\mathbf{u}$. Je-li $\mathbf{uG} \in \mathbb{F}[D^{-1}]$, je i $\mathbf{u} \in \mathbb{F}[D^{-1}]$ díky a).

„(4) $\Rightarrow$ (3)“:

a) Nechť je $\deg(\mathbf{uG}) = s < \infty$. Pak je $D^{-s}\mathbf{uG} \in \mathbb{F}[D^{-1}]$, a tedy také

$$D^{-s}\mathbf{u} = D^{-s}\mathbf{uG}\mathbf{G}'_{-1} \in \mathbb{F}[D^{-1}].$$

Z toho plyne, že $\deg(\mathbf{u}) \le s = \deg(\mathbf{uG})$. Je-li $\deg(\mathbf{uG}) = \infty$, je nerovnost zřejmá.

b) Nechť je nyní $\operatorname{del}(\mathbf{uG}) = s$. Pak je $D^{-s}\mathbf{uG} \in \mathbb{F}[[D]]$, a tedy také

$$D^{-s}\mathbf{u} = D^{-s}\mathbf{uG}\mathbf{G}' \in \mathbb{F}[[D]].$$

Z toho plyne, že $\operatorname{del}(\mathbf{u}) \ge s = \operatorname{del}(\mathbf{uG})$. Rovnost $\operatorname{del}(\mathbf{u}) = \operatorname{del}(\mathbf{uG})$ plyne z předpokladu realizovatelnosti $\mathbf{G}$.

□