Pokroky matematiky, fyziky a astronomie

František Katrnoška

Latinské čtverce a genetický kód

Pokroky matematiky, fyziky a astronomie, Vol. 52 (2007), No. 3, 177–187

Persistent URL: http://dml.cz/dmlcz/141355

Latinské čtverce a genetický kód

František Katrnoška, Praha

Letošního roku uplynulo již 300 let od narození Leonharda Eulera, kterého lze kromě jiného považovat za jednoho ze zakladatelů teorie grafů a teorie latinských čtverců. Tento článek je vyjádřením hluboké úcty k dílu tohoto geniálního matematika.

1. Úvod — latinské čtverce a jejich aplikace

Pojem latinského čtverce zavedl Euler roku 1782. Byl jím definován takto: *Latinský čtverec řádu n je schéma o n řádcích a n sloupcích, kde v každém řádku a v každém sloupci se každý prvek nějakého souboru mohutnosti n objevuje právě jednou.* Prvky latinského čtverce mohou být různé objekty, např. hrací karty, šachové figurky, osoby či přirozená čísla. Proslulou úlohou, která stála u zrodu pojmu latinského čtverce, byla úloha o 36 důstojnících, kterou zformuloval Euler r. 1779 a publikoval v roce 1782 (viz [10]). Jde o úlohu tohoto znění:

Seřaďte 36 důstojníků šesti různých hodností ze šesti různých pluků do čtvercového útvaru složeného ze 6 řad po 6 důstojnících tak, aby v každé řadě a v každém zástupu byly zastoupeny všechny hodnosti a všechny pluky.

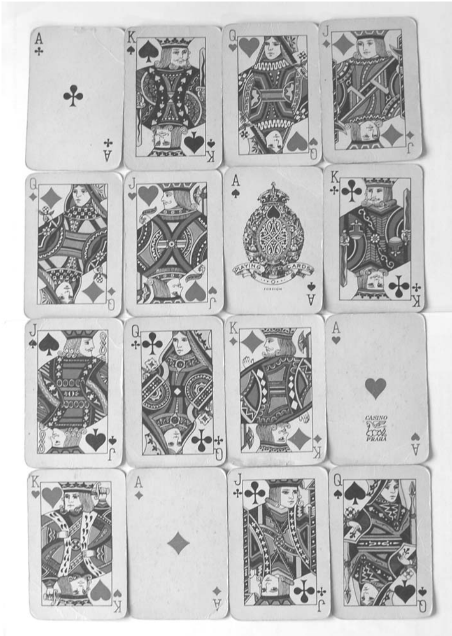
Euler se domníval, že tato úloha nemá řešení, avšak neuměl toto tvrzení dokázat. Trvalo pak dlouhých 118 let, než roku 1900 G. Tarry systematickým výpočtem dokázal (viz [29]), že tato úloha skutečně řešení nemá. Je třeba poznamenat, že mnoho prací věnovaných latinským čtvercům se týkalo úloh souvisejících s hazardními hrami. Například to byla úloha o seřazení 16 žolíkových karet předložená Jacquesem Ozanamem už roku 1694 a objevující se ve vydání [27] z roku 1723 (viz obr. 1).

V této souvislosti jsou velmi zajímavé problémy týkající se ortogonality latinských čtverců. Nechť jsou dány dva latinské čtverce L_1 a L_2, kde $L_1 = (a_{ij})$, $L_2 = (b_{ij})$, $i, j = 1, 2, \ldots, n$. Pak L_1 a L_2 nazýváme *ortogonální*, jestliže jsou všechny dvojice (a_{ij}, b_{ij}), $i, j = 1, 2, \ldots, n$, navzájem různé (viz [20], [25], [28]).

Zavedení pojmu ortogonality má svůj původ v Eulerově úloze o 36 důstojnících. Řešení této úlohy spočívalo v nalezení dvou vzájemně ortogonálních latinských čtverců šestého řádu. Jestliže jsou totiž latinské čtverce $L_1 = (a_{ij})$ a $L_2 = (b_{ij})$ ortogonální, pak v množině dvojic $E = (a_{ij}, b_{ij})$ se každá tato dvojice vyskytne právě jednou. Takový čtverec se pak nazývá *Eulerův čtverec*, viz [8], [21]. Snadno se zjistí, že neexistují ortogonální latinské čtverce řádu 2. Neexistence ortogonálních čtverců řádu 6 je vlastně

Doc. RNDr. František Katrnoška, CSc. (1929), Ústav matematiky VŠCHT, Technická 5, 166 28 Praha 6.

Předneseno na Semináři z dějin matematiky na Stavební fakultě ČVUT dne 24. dubna 2007.



Obr. 1. V každém řádku a v každém sloupci se nachází právě jedna z hodností: A, K, Q, J a zároveň právě jedna z barev: srdce, kára, piky, kříže (trefy).

zdůvodněním neřešitelnosti úlohy o 36 důstojnících. Konstrukcím Eulerových čtverců různých řádů (a tím i konstrukcím ortogonálních latinských čtverců) byla věnována celá řada prací (viz [3], [4], [8], [12], [21], [22], [23]).

L. Euler vyslovil domněnku, že když $n \equiv 2 \pmod 4$, pak neexistují vzájemně ortogonální čtverce řádu n. Ale hluboce se mýlil. Roku 1960 R. Bose, S. S. Shrikhande a E. T. Parker totiž v práci [5] dokázali, že ortogonální latinské čtverce existují pro všechna přirozená čísla n kromě případů $n = 2$ a $n = 6$. K danému latinskému čtverci však obecně nemusí existovat latinský čtverec k němu ortogonální, což dokázal H. B. Mann roku 1942 a zároveň stanovil nutnou podmínku pro existenci ortogonálního latinského čtverce (viz [23]). Celkově lze tedy říci, že v teorii ortogonality latinských čtverců jsou dva zásadní problémy, problém existence ortogonálního latinského čtverce k danému latinskému čtverci a s tím související konstrukce (viz [8], [12], [20], [22], [23], [28]). Hluboké souvislosti latinských čtverců s dalšími matematickými obory se ukázaly v uplynulých letech. Zmíníme se pouze o některých z nich.

Když se A. Cayley zabýval multiplikačními tabulkami konečných grup, zjistil, že tyto tabulky jsou latinskými čtverci. Na druhé straně latinský čtverec však nemusí být multiplikační tabulkou grupy. Později se ale, když R. Moufangová zavedla roku 1935 v práci [26] pojem kvazigrupy[1]), ukázalo, že každý latinský čtverec je multiplikační Cayleyovou tabulkou nějaké kvazigrupy (viz [1], [2], [8], [12]). Téhož roku Moufangová rovněž zjistila souvislost mezi kvazigrupami a nedesargueovskými projektivními geometriemi. Od té doby byly latinské čtverce používány jako prostředek ke stanovení

[1]) *Kvazigrupa* je neprázdná množina G s binární operací $\circ\colon G \times G \to G$ takovou, že pro každé $a, b \in G$ existuje právě jedno $x \in G$ a právě jedno $y \in G$, pro něž platí $a \circ x = b = y \circ a$.

existence konečných projektivních rovin různých řádů (viz [7], [20], [25], [28]). Proto se nyní o těchto rovinách, které připomínají klasické nekonečné projektivní roviny, stručně zmíníme.

Definice (viz [20]). *Konečná projektivní rovina řádu* $n \in \{1, 2, 3, \ldots\}$ je množina skládající se z $n^2 + n + 1$ bodů a z $n^2 + n + 1$ přímek, které splňují následující axiomy:

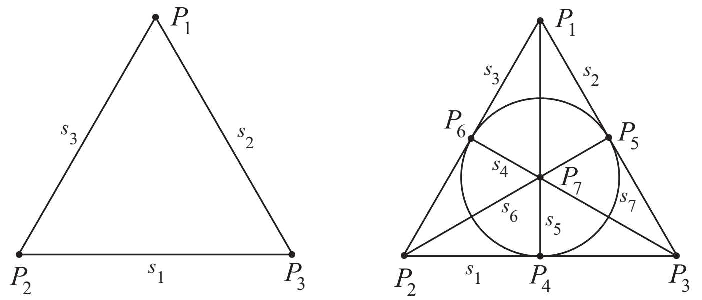
(A1) Každý bod leží na $n + 1$ přímkách.

(A2) Každá přímka obsahuje $n + 1$ bodů.

(A3) Libovolné dva různé body leží na právě jedné přímce.

(A4) Každé dvě různé přímky se protínají vždy jen v jednom bodě.

Jiné ekvivalentní definice konečné projektivní roviny jsou uvedeny např. v [12], [21], [25, str. 237], [28].



Obr. 2. Příklady konečných projektivních rovin pro $n = 1$ (trojúhelník) a pro $n = 2$ (Fanova rovina).

Předložený systém axiomů konečné projektivní roviny se ukazuje jako nejvhodnější, neboť výstižně charakterizuje dualitu mezi přímkami a body. Zaměníme-li totiž v systému axiomů (A1)–(A4) mezi sebou slova „bod“ a „přímka“ a zaměníme-li zároveň slova „leží“ a „protíná“, pak se systém axiomů nezmění.

Důsledkem toho je, že když zaměníme v pravdivém tvrzení týkajícím se projektivní roviny řádu n všude navzájem výše uvedené termíny, pak dostaneme opět pravdivý výrok. Snadno se zjistí, že nejmenší konečnou projektivní rovinou je trojúhelník (viz obr. 2). Méně triviálním příkladem projektivní roviny řádu $n = 2$ je tzv. Fanova rovina, která má podle definice 7 bodů a 7 přímek. Zajímavost této roviny spočívá v tom, že jednu z jejích přímek je nutno zobrazit jako kružnici (viz obr. 2). Ostatní přímky se zobrazují jako úsečky. Tento příklad konečné projektivní roviny řádu 2 uvedl O. Veblen v [31], kde zároveň ukázal, že tato rovina nemůže být graficky znázorněna pouze s použitím úseček.

Ukázalo se, že existence konečné projektivní roviny řádu n úzce souvisí s existencí $n - 1$ vzájemně ortogonálních čtverců řádu n. Poznamenejme ještě, že existuje-li $n - 1$ vzájemně ortogonálních latinských čtverců řádu n, pak tento systém čtverců se nazývá *úplný*. Uveďme nyní následující větu, jejíž důkaz lze nalézt v [28]:

Věta 1. *Nechť* $n \geqq 2$. *Projektivní rovina řádu* n *existuje tehdy a jen tehdy, když existuje úplná množina* $n-1$ *vzájemně ortogonálních latinských čtverců řádu* n.

V roce 1938 R. C. Bose v práci [4] ukázal, proč neexistují projektivní roviny řádu 6. Vzhledem k tomu, že neexistují vzájemně ortogonální latinské čtverce řádu 6, plyne rovněž z věty 1 neexistence konečné projektivní roviny řádu 6. Neexistence konečné projektivní roviny řádu 10 byla dokázána (viz [20]) neobyčejně rozsáhlými výpočty na počítačích v USA. Pro existenci projektivní roviny řádu n platí následující věta (viz [7], [12], [28]):

Věta 2. *Konečná projektivní rovina řádu* n *existuje, jestliže* n *je mocninou prvočísla.*

Je známo, že pro $n = 2$, 3, 4, 5, 7 a 8 je konečná projektivní rovina určena jednoznačně. Pro $n = 9$ ale existují 4 vzájemně neizomorfní konečné projektivní roviny (viz [20]). Poznamenejme ještě, že v roce 1949 byla dokázána následující věta (viz [6]).

Věta 3 (Bruckova-Ryserova). *Jestliže* $n \equiv 1, 2 \pmod 4$, *pak nutná podmínka pro existenci konečné projektivní roviny řádu* n *je, že existují celá čísla* x *a* y *tak, že* $n = x^2 + y^2$.

Tato podmínka není ale postačující, neboť $10 \equiv 2 \pmod 4$ a $10 = 1^2 + 3^2$. Bruckova-Ryserova věta také nic neříká o případu $n = 12$, který zatím není rozřešen. Na druhé straně okamžitě implikuje neexistenci konečné projektivní roviny například pro $n = 14$, neboť $14 \equiv 2 \pmod 4$ a číslo 14 nelze vyjádřit jako součet dvou čtverců.[2])

K řešení některých problémů týkajících se latinských čtverců byla použita též teorie grafů. K tomuto účelu byl zaveden pojem grafu latinského čtverce (viz [8]):

Nechť L je latinský čtverec řádu n, kde $L = (a_{ij})$, $i, j = 1, \dots, n$, $a_{ij} \in X$, $\operatorname{card} X = n$. *Grafem latinského čtverce* se nazývá graf $G_L = (V, E)$, jehož vrcholy jsou tvaru $v_{ij} = (a_{i1}, a_{1j}, a_{ij})$ a jehož hrany jsou spojnice těch vrcholů G_L, které mají alespoň jednu ze svých souřadnic stejnou.

Roku 1970 použili F. C. Bussemaker a J. J. Seidel grafy dvanácti hlavních tříd latinských čtverců řádu 6 k sestrojení regulárních symetrických Hadamardových matic[3]) řádu 36. Je třeba dodat, že Hadamardovy matice se používají mj. ke konstrukci některých blokových schémat (viz [12], [28]).

Latinské čtverce lze též interpretovat jako úplné bichromatické (bipartitní) grafy (viz [13], [25], [28]), což jsou úplné grafy tvaru $G = (V_1, V_2, E)$ takové, že množina vrcholů V je rovna $V = V_1 \cup V_2$, přičemž $V_1 \cap V_2 = \emptyset$ a množina hran E se skládá z hran tvaru $h = (v_1, v_2)$, kde $v_1 \in V_1$ a $v_2 \in V_2$. Bichromatické grafy odpovídající latinským čtvercům se používají pro znázornění konečných projektivních rovin (viz [21], [28], [33]).

---

[2]) Poznamenejme, že prvočísla tvaru $n \equiv 1 \pmod 4$ lze jednoznačně rozložit na součet dvou čtverců, jak již dokázal Pierre de Fermat (viz [19]).

[3]) *Hadamardovou maticí* řádu n nazýváme matici H typu $n \times n$, jejímiž prvky jsou pouze čísla 1 a -1, pro niž $HH^\top = nI$, kde I je jednotková matice.

Závěrem této kapitoly je možno poukázat na to, že s aplikacemi latinských čtverců se lze setkat při kódování tajných dokumentů, při plánování experimentů v operačním výzkumu (viz [11]) nebo při pěstování několika odrůd zemědělských plodin na menších čtvercových políčkách o různé bonitě půdy (viz [3]). Latinské čtverce se používají i při organizaci golfových, bridgeových či tenisových turnajů [22]. Jako kuriózní příklad latinského čtverce řádu 9 je též každé výsledné postavení populární hry sudoku (viz [14]).

2. Zobecněné latinské čtverce

Jako historicky první zobecněné latinské čtverce byly zavedeny latinské obdélníky (viz [8], [12], [25], [28]), latinské čtverce vzhledem k řádkům (angl. row latin squares), nekonečné latinské čtverce a různé další druhy neúplných latinských čtverců (viz [22]). V pozdějších letech K. Kishen a nezávisle R. A. Fisher zavedli latinské krychle a hyperkrychle (1942–1945). Pro obšírnost této problematiky se nebudeme těmito druhy zobecněných latinských čtverců zabývat. Podrobné informace o nich najde čtenář v knize [8] autorů J. Denese a A. D. Keedwella.

Symbol R bude dále označovat komutativní okruh s jednotkovým prvkem e. Zavedeme nyní zobecněné latinské čtverce nad R (viz [16]).

Definice. *Zobecněným R-latinským čtvercem řádu n* nazveme čtvercovou matici L řádu n nad R takovou, že součty všech prvků každého řádku a každého sloupce jsou identické a jsou rovny témuž prvku $s \in R$.

Poznamenejme, že zobecněné R-latinské čtverce mohou být ekvivalentně zavedeny též takto: Čtvercová matice L řádu n nad R je zobecněný R-latinský čtverec, jestliže matice L i $L^\top$ mají tentýž vlastní vektor $(e, e, \dots, e)^\top$ odpovídající stejné vlastní hodnotě $\lambda \in R$.

Množinu všech zobecněných R-latinských čtverců řádu n označíme symbolem $\mathcal{L}_n(R)$. Jak bude ukázáno v této kapitole, dosud známé latinské čtverce, magické čtverce[4]), permutační matice[5]), dvojitě stochastické matice[6]) jsou speciálními případy zobecněných R-latinských čtverců. Z tohoto důvodu je zavedení zobecněných R-latinských čtverců účelné a vhodné.

Pro $L \in \mathcal{L}_n(R)$ definujme zobrazení $w\colon \mathcal{L}_n(R) \to R$ vztahem $w(L) = s$, kde s je součet prvků libovolného řádku a také libovolného sloupce matice L. Zobrazení w budeme nazývat *váhou*. Například pro hru sudoku 9×9 je $s = 45$.

---

[4]) *Magickým čtvercem řádu n* rozumíme čtvercovou matici řádu n, jejímiž prvky je n^2 celých nezáporných čísel (nikoliv nutně po sobě jdoucích ani nutně různých) tak, že součet čísel v libovolném řádku, v libovolném sloupci i v obou úhlopříčkách je týž (srov. obr. 3).

[5]) Permutační matice má v každém řádku a každém sloupci právě jeden jednotkový prvek $e \in R$, ostatní prvky jsou nulové.

[6]) Reálná matice se nazývá *dvojitě stochastická*, jestliže všechny její prvky jsou nezáporné a součet prvků v každém jejím řádku a sloupci je roven 1.

$$\begin{bmatrix} 2 & 7 & 6 \\ 9 & 5 & 1 \\ 4 & 3 & 8 \end{bmatrix}$$

Obr. 3. Magický čtverec 3×3. Je příkladem zobecněného $\mathbb{Z}$-latinského čtverce, který není latinským čtvercem.

Uveďme nyní několik příkladů zobecněných R-latinských čtverců. Snadno lze ukázat, že každý latinský čtverec řádu n nad R je prvkem $\mathcal{L}_n(R)$. Symbol $\mathbb{Z}$ bude v dalším textu vždy označovat okruh všech celých čísel a symbol $\mathbb{R}$ těleso reálných čísel. Magický čtverec řádu n splňuje podmínky předchozí definice, a tedy je prvkem $\mathcal{L}_n(\mathbb{Z})$. Jestliže P je permutační matice řádu n nad R, pak $P \in \mathcal{L}_n(R)$ a $w(P) = e$. V případě, že D je dvojitě stochastická matice řádu n, platí $D \in \mathcal{L}_n(\mathbb{R})$ a $w(D) = 1$.

Uveďme ještě jako příklad (viz [24, str. 19]) zobecněného $\mathbb{Z}_2$-latinského čtverce pro

$$\mathbb{Z}_2 = \mathbb{Z}/\bmod 2 = \{0, 1\}$$

incidenční matici konečné projektivní roviny řádu n, což je čtvercová matice (a_{ij}) řádu $n^2 + n + 1$, pro kterou platí: jestliže bod P_i této roviny leží na přímce s_j, pak $a_{ij} = 1$, v opačném případě klademe $a_{ij} = 0$ pro $i, j = 1, \dots, n$. Následující matice je incidenční matice Fanovy roviny řádu 2 (srov. obr. 2):

$$\begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}.$$

Algebraickou strukturu množiny $\mathcal{L}_n(R)$ lze popsat následující větou (viz [17]).

Věta 4. *Množina $\mathcal{L}_n(R)$ je vzhledem k obvyklým maticovým operacím involutivní*[7]) *okruh, který je pro $n > 2$ nekomutativní. Přitom platí, že*

$$w(L_1 L_2) = w(L_2) w(L_1) = w(L_1) w(L_2)$$

a $w(L^*) = w(L)$ pro $L_1, L_2, L \in \mathcal{L}_n(R)$.

Snadno nahlédneme, že váha w je homomorfismus okruhu $\mathcal{L}_n(R)$ na R a že $w^{-1}(0)$ je oboustranný ideál okruhu $\mathcal{L}_n(R)$ (viz [24, str. 129]). Přitom platí následující věta.

[7]) Okruh se nazývá *involutivní*, jestliže je v něm definováno bijektivní zobrazení $*\colon\ R \to R$ takové, že $(x^*)^* = x$, $(x+y)^* = x^* + y^*$ a $(xy)^* = y^* x^*$ pro všechna $x, y \in R$. Jestliže má navíc R jednotkový prvek e, pak ještě požadujeme, aby $e^* = e$.

Věta 5. *Faktorokruh* $\mathcal{L}_n(R)/\operatorname{Ker} w$ *je izomorfní s okruhem* R.

Další algebraické vlastnosti $\mathcal{L}_n(R)$ jsou uvedeny v práci [17], která je věnována R-algebrám zobecněných R-latinských čtverců.

Na závěr této kapitoly uveďme ještě některé vlastnosti váhy w. Symbolem $\mathcal{S}_n(R)$ označme množinu všech permutačních matic řádu n nad okruhem R. Pak zřejmě platí $\mathcal{S}_n(R) \subset \mathcal{L}_n(R)$, a jestliže $P \in \mathcal{S}_n(R)$, pak $w(P) = e$. Zaveďme nyní bijektivní zobrazení f_P^ℓ, f_P^r a f_T, která zobrazují okruh $\mathcal{L}_n(R)$ na sebe následujícím způsobem:

Jestliže $P \in \mathcal{S}_n(R)$ a $L \in \mathcal{L}_n(R)$, pak položíme $f_P^\ell = P \cdot L$, $f_P^r = L \cdot P$ a $f_T = L^\top$. Zobrazení f_P^ℓ, resp. f_P^r jsou vlastně permutace řádků, resp. sloupců zobecněného R-latinského čtverce a zobrazení f_T je involucí okruhu $\mathcal{L}_n(R)$. Užitím věty 4 dostáváme, že

$$w\big(f_P^\ell(L)\big) = w(P \cdot L) = w(P)w(L) = e\,w(L) = w(L).$$

Analogicky platí $w\big(f_P^r(L)\big) = w(L)$ a $w\big(f_T(L)\big) = w(L^\top) = w(L)$. Odtud pak plyne, že bijektivní zobrazení f_P^ℓ, f_P^r a f_T zachovávají stejnou hodnotu váhy zobecněných R-latinských čtverců.

3. Zobecněné latinské čtverce v genetice

V tomto odstavci uvedeme aplikace zobecněných R-latinských čtverců v genetice. Pro orientaci čtenáře se nejprve zmíníme o některých pojmech a výsledcích dosažených v matematické genetice. Je všeobecně známo, že proteiny (bílkoviny) jsou základními stavebními kameny všech živých organismů. Je rovněž známo, že se proteiny skládají z aminokyselin, které jsou mezi sebou vzájemně spojeny peptidickými vazbami, a že zpravidla vytvářejí lineární řetězce. Ze všech aminokyselin je pouze 20 těch, které se účastní při stavbě proteinů. Jejich syntéza je řízena procesem zvaným translace. Pro zavedení tohoto pojmu označme symbolem K množinu

$$K = \{A, U, C, G\},$$

kde A je nukleová báze adenin, U uracyl, C cytosin a G guanin (viz [18]). Tyto nukleotidy vytvářejí vzájemným spojováním trojice zvané *triplety* a spolu s cukrem ribózou a kyselinou fosforečnou vytvářejí kyselinu ribonukleovou RNA. Prostřednictvím tripletů nukleových bází A, U, C, G jsou kódovány odpovídající aminokyseliny. Označme dále symbolem $\mathcal{A}$ množinu zmíněných 20 aminokyselin a tří slov amber, ochre a opal, což jsou terminační symboly znamenající ukončení syntézy proteinu (a také jména nerostů). Za včlenění určité aminokyseliny do proteinového řetězce odpovídá vždy pouze trojice nukleotidů. *Translací* budeme tedy rozumět zobrazení $t\colon K \times K \times K \to \mathcal{A}$. Každý triplet bází kóduje buď určitou aminokyselinu, nebo označuje počátek či ukončení syntézy proteinu.

Zdůvodnění faktu, proč k určení konkrétní aminokyseliny z množiny $\mathcal{A}$ jsou vhodné právě jen 4 nukleové báze, je uvedeno v práci [9]. Translace je nejdůležitější fáze procesu kódování. Definitivní tvar genetického kódu byl stanoven r. 1966 M. W. Nirenbergem sestavením tabulky tripletů pro všechny bílkovinotvorné aminokyseliny.

V tomto článku můžeme tuto tabulku kódů uvést poněkud pozměněnou, a to ve tvaru čtvercové matice řádu 8.

Tabulka 1. Všechny možné triplety nukleotidů z množiny K. Každý z nich kóduje některou bílkovinotvornou aminokyselinu nebo označuje ukončení syntézy bílkoviny.

$$\begin{bmatrix} UUU & UCU & UAU & UGU & | & CUU & CCU & CAU & CGU \\ UUC & UCC & UAC & UGC & | & CUC & CCC & CAC & CGC \\ UUA & UCA & UAA & UGA & | & CUA & CCA & CAA & CGA \\ UUG & UCG & UAG & UGG & | & CUG & CCG & CAG & CGG \\ --- & --- & --- & --- & | & --- & --- & --- & --- \\ AUU & ACU & AAU & AGU & | & GUU & GCU & GAU & GGU \\ AUC & ACC & AAC & AGC & | & GUC & GCC & GAC & GGC \\ AUA & ACA & AAA & AGA & | & GUA & GCA & GAA & GGA \\ AUG & ACG & AAG & AGG & | & GUG & GCG & GAG & GGG \end{bmatrix}$$

Příklad 1. Nyní ukážeme, že tabulku 1 lze ekvivalentně zapsat pomocí zobecněného R-latinského čtverce. Za tímto účelem si zvolme cyklickou grupu $\mathcal{G}$ řádu 4, kde $\mathcal{G} = \{e, g, g^2, g^3\}$. Počet prvků $\mathcal{G}$ přitom odpovídá počtu nukleových bází. Označme nyní symbolem $\mathcal{R}_\mathcal{G}$ komutativní okruh nad $\mathbb{Z}$, jehož množina generátorů je grupa $\mathcal{G}$. Okruh $\mathcal{R}_\mathcal{G}$ je izomorfní Gaussovu oboru integrity $\mathbb{Z}(i)$ (viz [24]). Konstrukci zobecněného $\mathcal{R}_\mathcal{G}$-latinského čtverce odpovídajícího tabulce 1 provedeme nyní takto. V tabulce 1 položíme

$$U = e, \quad C = g, \quad A = g^2, \quad G = g^3$$

a tyto prvky mezi sebou vynásobíme. Jestliže použijeme základní vlastnosti grupy $\mathcal{G}$, pak obdržíme následující $\mathcal{R}_\mathcal{G}$-latinský čtverec řádu 8:

$$L = \begin{bmatrix} e & g & g^2 & g^3 & | & g & g^2 & g^3 & e \\ g & g^2 & g^3 & e & | & g^2 & g^3 & e & g \\ g^2 & g^3 & e & g & | & g^3 & e & g & g^2 \\ g^3 & e & g & g^2 & | & e & g & g^2 & g^3 \\ -- & -- & -- & -- & | & -- & -- & -- & -- \\ g^2 & g^3 & e & g & | & g^3 & e & g & g^2 \\ g^3 & e & g & g^2 & | & e & g & g^2 & g^3 \\ e & g & g^2 & g^3 & | & g & g^2 & g^3 & e \\ g & g^2 & g^3 & e & | & g^2 & g^3 & e & g \end{bmatrix}.$$

Označíme-li symbolem B matici

$$B = \begin{bmatrix} e & g & g^2 & g^3 \\ g & g^2 & g^3 & e \\ g^2 & g^3 & e & g \\ g^3 & e & g & g^2 \end{bmatrix},$$

pak můžeme matici L zapsat v blokovém tvaru

$$L = \begin{bmatrix} B & gB \\ g^2B & g^3B \end{bmatrix}.$$

Snadno se lze přesvědčit, že samotné bloky B, gB, g^2B a g^3B řádu 4 jsou také latinské čtverce nad $\mathcal{R}_\mathcal{G}$. Připomeňme ještě, že v tabulce 1 představující genetické kódy nejsou uvedeny kódované aminokyseliny (vzhledem k tomu, že konstrukce zobecněného latinského čtverce je na tom nezávislá). Pro podrobnější informaci o genetickém kódování doporučujeme čtenáři práce [18], [30] a [32].

Na závěr tohoto příkladu uveďme ještě jednu zajímavost. Jestliže v tabulce 1 vynecháme ve všech tripletech první (nebo druhou) nukleovou bázi, pak dostaneme blokovou matici řádu 2 následujícího tvaru:

$$\mathcal{D} = \begin{bmatrix} D & D \\ D & D \end{bmatrix}, \quad \text{kde} \quad D = \begin{bmatrix} UU & CU & AU & GU \\ UC & CC & AC & GC \\ UA & CA & AA & GA \\ UG & CG & AG & GG \end{bmatrix}.$$

Provedeme-li nyní v matici D záměnu nukleových bází za prvky grupy $\mathcal{G}$ (již uvedeným způsobem), pak výslednou maticí bude matice B.

Z tabulky genetického kódu (viz [18, str. 213]) je možno zjistit, že většina aminokyselin je kódována více triplety. Například alanin je kódován čtyřmi různými triplety GCA, GCU, GCC a GCG lišícími se jen třetí nukleovou bází. Arginin, leucin a serin jsou kódovány dokonce šesti různými triplety. Častým jevem, který může postihnout genetický kód, je tzv. mutace, při níž dochází ke změně nukleové báze některého tripletu, k jeho ztrátě (deleci) nebo ke včlenění nové skupiny bází do procesu kódování. Bodová mutace spočívá ve změně právě jedné nukleové báze tripletu.

Příklad 2. Zaveďme nejprve následující označení. Nechť p_r je pravděpodobnost, že nedojde k žádné záměně nukleových bází. Označme p_1, resp. p_2 pravděpodobnost, že dojde k záměně nukleových bází $U \leftrightarrow C$, resp. $A \leftrightarrow G$. (Podobně lze uvažovat záměnu $U \leftrightarrow G$, resp. $A \leftrightarrow C$; nebo $U \leftrightarrow A$, resp. $C \leftrightarrow G$.) Experimentálně bylo zjištěno, že při všech těchto záměnách nukleových bází tripletů platí:

(i) $0 < p_2 < p_1 \ll p_r < 1$.

Vzhledem k tomu, že pro každou bázi existuje vždy jedna záměna, která je nejpravděpodobnější (zatímco ostatní dvě jsou méně pravděpodobné), a mimo to báze může zůstat i nezměněna, plyne odtud, že

(ii) $p_1 + 2p_2 + p_r = 1$.

Zdůrazněme, že všechny tyto námi uvedené pravděpodobnosti se týkají všech bodových mutací, nikoliv pouze na třetí bázi tripletu. Následující symetrická dvojitě stochastická matice charakterizuje tuto bodovou mutaci

$$M = \begin{bmatrix} p_r & p_1 & p_2 & p_2 \\ p_1 & p_r & p_2 & p_2 \\ p_2 & p_2 & p_r & p_1 \\ p_2 & p_2 & p_1 & p_r \end{bmatrix}.$$

Vzhledem k (i) a (ii) odtud plyne, že M je zobecněný $\mathcal{R}_{\mathcal{G}}$-latinský čtverec řádu 4. Označme nyní

$$M_1 = \begin{bmatrix} p_r & p_1 \\ p_1 & p_r \end{bmatrix}, \quad M_2 = \begin{bmatrix} p_2 & p_2 \\ p_2 & p_2 \end{bmatrix}.$$

Vidíme, že matice M_1 je latinský čtverec, zatímco M_2 je zobecněný $\mathcal{R}_{\mathcal{G}}$-latinský čtverec. Matici M lze vyjádřit blokově ve tvaru

$$M = \begin{bmatrix} M_1 & M_2 \\ M_2 & M_1 \end{bmatrix}.$$

Je opět zřejmé, že bloková matice M je latinský čtverec řádu 2. Pro podrobnější informaci viz [15].

Na závěr ještě poznamenejme, že zobecněné latinské čtverce lze zavést též nad komutativními polookruhy (viz [17]). Některé výsledky pro takové zobecněné čtverce rovněž platí.

Poděkování. Autor děkuje RNDr. A. Šolcové, Ph. D., a prof. L. Somerovi z Washingtonu, D. C., za velmi cenné připomínky a Mgr. Ing. J. Šolcovi za pomoc s obrázky. Podpořeno grantem MSM 6046137306 (Ústav matematiky VŠCHT).

Literatura

[1] Albert, A. A.: *Quasigroups I, II.* Trans. Amer. Math. Soc. *54* (1943), 507–519; *55* (1944), 401–419.

[2] Belousov, V. D.: *Osnovy teorii kvazigrup i lup.* Izd. Nauka, Moskva 1967.

[3] Bosák, J.: *Latinské štvorce.* Škola mladých matematiků, sv. 38, Praha 1976.

[4] Bose, R. C.: *On the application of the properties of Galois fields to the problem of construction of hyper-Graeco-Latin squares.* Sankya *3* (1938), 323–338.

[5] Bose, R. C., Parker, E. T., Shrikhande, S. S.: *Further results on the construction of mutually orthogonal Latin squares and the falsity of Euler's conjecture.* Canad. Math. J. *12* (1960), 189–203.

[6] Bruck, R. H., Ryser, H. J.: *On the non-existence of certain finite projective planes.* Canad. Math. J. *1* (1949), 88–93.

[7] Dembowski, P.: *Finite geometries.* Springer, Berlin 1968.

[8] Dénes, J., Keedwell, A. D.: *Latin squares and their applications.* Akadémiai Kiadó, Budapest, Academic Press, London 1974.

[9] Deng, Bo: *Why is the number of DNA bases 4?* Bull. Math. Biol. *68* (2006), 727–733.

[10] Euler, L.: *Recherches sur une nouvelle espèce de quarrés magiques.* Verh. Zeeuwsch. Genootsch. Wetensch. Vlissengen *9* (1782), 85–239.

[11] Fisher, R. A.: *The design of experiments* (8th edition). Olivier et Boyd, Edinburgh 1966.

[12] Hall, M.: *Combinatorial theory.* Blaisdell Publ. Comp., Toronto, London 1967.

[13] Harary, F.: *Graph theory.* Addison-Wesley, Reading Mass 1969.

[14] Herzberg, A. L., Murty, M. R.: *Sudoku squares and chromatic polynomials.* Notices Amer. Math. Soc. *54* (2007), 708–717.

[15] Kárná, L.: *Genetic code from the point of view of code theory.* Proc. ICPM '05, Liberec 2005, 267–273.
[16] Katrnoška, F.: *Logics that are generated by idempotents.* Lobachevskii J. Math. *15* (2004), 11–19.
[17] Katrnoška, F.: *On algebras of generalized R-Latin squares.* Připraveno pro tisk.
[18] Katrnoška, F., Křížek, M.: *Genetický kód a teorie monoidů aneb 50 let od objevu struktury DNA.* PMFA *48* (2003), 207–222.
[19] Křížek, M., Luca, F., Somer, L.: *17 lectures on Fermat numbers: From number theory to geometry.* Springer, New York 2001.
[20] Lam, C. W. H.: *The search for a finite projective plane of order 10.* Amer. Math. Monthly *98* (1991), 305–318.
[21] Laugwitz, D.: *Eulerovy čtverce.* PMFA *25* (1980), 69–79.
[22] Laywine, Ch., Mullen, G. L.: *Discrete mathematics using Latin squares.* Wiley Interscience 1998.
[23] Mann, H. B.: *The construction of orthogonal Latin squares.* Ann. Math. Stat. *13* (1942), 418–423.
[24] Markus, M.: *Introduction to modern algebra.* Marcel Dekker, New York, Basel 1978.
[25] Matoušek, J., Nešetřil, J.: *Kapitoly z diskrétní matematiky.* Karolinum, Praha 2000.
[26] Moufang, R.: *Zur Struktur von alternativ Körpern.* Math. Ann. *110* (1935), 416–430.
[27] Ozanam, J.: *Récréations mathématiques et physiques 1723* (angl. vyd. 1814, Hutton).
[28] Rybnikov, K. A.: *Vvedënije v kombinatornyj analiz.* Izd. Moskovskogo Univ., Moskva 1985.
[29] Tarry, G.: *Le problème des 36 officiers.* C. R. Assoc. Fr. Av. Sci. *1* (1900), 122–123; *2* (1901), 170–203.
[30] Thompson, J. S., Thompson, M. W.: *Genetics in medicine.* W. B. Saunders Company, Philadelphia 2001.
[31] Veblen, O.: *A system of axioms for geometry.* Trans. Amer. Math. Soc. *5* (1904), 343–384.
[32] Watson, J. D.: *Molekulární biologie genu.* Academia, Praha 1982.
[33] Wilson, R. J.: *Introduction to graph theory.* Oliver and Boyd, Edinburgh 1972.