INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Předmluva ke druhému číslu časopisu Kvaternion

Milí čtenáři, právě držíte v rukou druhé číslo časopisu Kvaternion, který v současné době slouží především jakožto publikační platforma sítě pracovišť projektu AMathNet. Jak již bylo uvedeno v úvodníku prvního čísla, nejedná se o časopis ryze vědecký, jeho smyslem je transfer znalostí zejména v oblasti aplikací matematiky, publikace článků popularizačního a didaktického charakteru se snahou o maximální zapojení studentů jednotlivých partnerů projektu. Tento charakter je beze zbytku naplněn druhým číslem, které vám s potěšením představujeme. Druhé číslo je obsahově složeno zejména ze čtyř příspěvků 2. workshopu "Aplikovaná matematika" pořádaného FSI VUT v Brně ve dnech 31.1–3.2. 2012 v Ostravici: „Replikátorová rovnice – příklad užití matematiky v biologii" (autor Zdeněk Pospíšil), „Šifrovací metoda založená na fraktální kompresi" (autor Tomáš Grísa), „Weilovo párování na eliptických křivkách v kryptosystémech založených na totožnosti a kryptograficky randomizované odpovídací techniky" (autor Miroslav Kureš) a „Od kompozitních materiálů ke slabé konvergenci" (autor Jan Franců). Příspěvek „O kvaternionových algebrách" (autor Lenka Macálková) vznikl svým způsobem také v Ostravici na základě diskusí studentky doktorského studia L. Macálkové s doc. Kurešem. Příspěvek s názvem „Constructed Quaternary Cryptographic Functions Class" (autoři Zoubida Jadda a Patrice Parraud) souvisí s návštěvou dr. Zoubidy Jaddy na FSI VUT v Brně podpořenou projektem AMathNet. Věříme, že si každý z čtenářů v právě představeném čísle najde oblast, která jej obohatí o nové poznatky v oblasti aplikací matematiky a která bude impulsem pro další rozvoj jeho odborné činnosti.

Olomouc, 12.8. 2012

Radomír Halaš
t.č. vedoucí katedry Algebry a Geometrie
PřF UP Olomouc

CONSTRUCTED QUATERNARY CRYPTOGRAPHIC FUNCTIONS CLASS

ZOUBIDA JADDA AND PATRICE PARRAUD

ABSTRACT. New results on quaternary ($\mathbb{Z}_4 = \{0, 1, 2, 3\}$-valued) cryptographic functions are presented. We define and characterize completely the $\mathbb{Z}_4$-balancedness and the $\mathbb{Z}_4$-nonlinearity according to the HAMMING metric and the LEE metric. In the particular case of quaternary Bent functions we show that the maximal nonlinearity of these functions is bounded for the HAMMING metric and we give the exact value of the maximal nonlinearity of these functions for the LEE metric. A general construction, based on Galois ring, is related in detail and applied to obtain a class of balanced and high nonlinearity quaternary cryptographic functions. We use Gray map to derive these constructed quaternary functions to obtain balanced Boolean functions having high nonlinearity.

1. INTRODUCTION

Boolean ($\{0, 1\}$-valued) functions of length n used in pseudo-random generators of stream and block ciphers play an important role in their security ([7, 1]). These functions are usually studied over finite field of two elements $\mathbb{F}_2$. Finding Boolean functions with optimal cryptographic properties as balancedness and high nonlinearity is still an open problem. The purpose of this paper is to present new results on quaternary ($\{0, 1, 2, 3\}$-valued) cryptographic functions. This work is motivated by the interest in studying quaternary objects and structures (see [8, 12]). The usual metric used in $\mathbb{Z}_4$ is the LEE metric which allows to have an isometry from ($\mathbb{Z}_4^m$, LEE distance) to ($\mathbb{F}_2^{2m}$, HAMMING distance) with the Gray map. We begin by defining and characterizing exactly quaternary cryptographic functions of length m. Then, we formally describe balancedness and nonlinearity over $\mathbb{Z}_4$ according to the HAMMING metric and the LEE metric. Quaternary Bent functions [19] (or more generaly q-ary Bent functions [3, 9, 10, 11]) are defined by Walsh transform. For m-variables quaternary Bent functions we prove that the maximal nonlinearity is bounded between $3 \cdot 4^{m-1} - 2^{m-1}$ and $3 \cdot 4^{m-1} - 2^{m-2}$ under the HAMMING metric and we give conditions to reach the upper bound. We show that the exact value of the maximal nonlinearity of these functions under the LEE metric is $4^{m-1} - 2^{m+1}$. A general construction of quaternary cryptographic functions is detailed, using cyclotomic classes of the multiplicative group of a Galois ring R. We point out the fact that the balancedness and the nonlinearity of the obtained

Keywords. Quaternary functions, Boolean functions, cryptographic functions, nonlinearity, balancedness, quaternary algebra, Gray map, Galois ring.

The work was supported by the project A-Math-Net – Applied Mathematics Knowledge Transfer Network (CZ.1.07/2.4.00/17.0100).

functions depend on the b-polynomial used to construct R and on the distribution of these classes over R. We naturally apply this construction to a particular configuration in order to obtain a class of m-variables quaternary cryptographic functions which are balanced and have nonlinearity bounded between $3\cdot 4^{m-1}-2^m$ and $3\cdot 4^{m-1}-2^{m-1}$ for the HAMMING metric and bounded between 4^m-2^{m+1} and 4^m-2^m for the LEE metric. Using the Gray map with these obtained quaternary functions we present $2m-$variables balanced Boolean functions with high nonlinearity. To avoid any confusion, an n-variables Boolean function is denoted by f while an m-variables quaternary function is denoted by F .

2. Boolean Functions Basics

Let n be a natural integer and $\mathbb{F}_2^n$ the set of all n-tuples of elements in the finite field $\mathbb{F}_2=\{0,1\}$ with its sum denoted by $\oplus$. An n-variables Boolean function f is a function from $\mathbb{F}_2^n$ to $\mathbb{F}_2$ which can be identified by its truth table $[f(0,\cdots,0),\cdots,f(1,\cdots,1)]$ of length 2^n. The support of f is defined by $supp(f)=\{u\in\mathbb{F}_2^n \mid f(u)\neq 0\}$ and the Hamming weight $w_H(f)$ of f by the size of its support. The Hamming distance between two n-variables Boolean functions f and g is $d_H(f,g)=w_H(f\oplus g)$ where $\oplus$ denotes the addition in $\mathbb{F}_2$. The Walsh transform of an n-variables Boolean function f is the complex mapping from $\mathbb{F}_2^n$ to $\mathbb{C}$ defined by $W_f(u)=\sum_{v\in\mathbb{F}_2^n}(-1)^{u\cdot v+f(v)}$ where $u\cdot v$ denotes the usual inner product in $\mathbb{F}_2^n$. An n-variables Boolean function f is balanced if its truth table contains an equal number of 1's and 0's which means that $w_H(f)=2^{n-1}$ or in spectral term $W_f(0)=0$. The nonlinearity of a n-variables Boolean function f is the minimum distance to all affine functions $nl_2(f)=\min\limits_{g\ affine} d_H(f,g)$. Using the Walsh transform, the nonlinearity of f can be expressed by $nl_2(f)=2^{n-1}-\frac{1}{2}\max\limits_{a\in\mathbb{F}_2^n}|W_f(a)|$. Readers can read [5, 6, 16] for more detailed explanations of Boolean functions cryptographic criteria. For every n-variables Boolean function f, we have $nl_2(f)\leq 2^{n-1}-2^{\frac{n}{2}-1}$. This bound is reached for Bent functions [17, 14] which are characterised by $\forall u\in\mathbb{F}_2^n,\quad |W_f(u)|=2^{\frac{n}{2}}$ for n even. A Bent function could not be balanced. Finding maximal nonlinearity Boolean functions (see [13, 15, 18]) is an open problem.

3. Quaternary Cryptographic Functions

3.1. Quaternary Tools

Throughout this section, i will denote the complex number such that $i^2=-1$. Let $\mathbb{Z}_4=\mathbb{Z}/4\mathbb{Z}=\{0,1,2,3\}$ be the ring of integers modulo 4 which is group-isomorphic to $\mathbb{U}_4=\{\pm 1,\pm i\}$ the group of 4^{th} root of unity in $\mathbb{C}$ under the standard isomorphism $x\to i^x$. $\mathbb{Z}_4^m$ will represent the set of all $m-$tuples of elements in $\mathbb{Z}_4$ where m is a natural integer. The addition on $\mathbb{Z}_4$ (addition (mod 4)) will be denoted by $+$. The LEE weights w_L of $0,1,2,3$ in $\mathbb{Z}_4$ are $0,1,2,1$ respectively and the LEE weight $w_L(u)$ of an element u of $\mathbb{Z}_4^m$ is the rational sum of the LEE weight of its components. The LEE distance $d_L(u,v)$ between two elements u and v in $\mathbb{Z}_4^m$ is $w_L(u+v)$.

Definition 3.1. An m-variables quaternary function F is a function from $\mathbb{Z}_4^m$ to $\mathbb{Z}_4$ which can be identified by its truth table $[F(0,\cdots,0),\cdots,F(3,\cdots,3)]$ of length 4^m. Let us define $\mathcal{F}(\mathbb{Z}_4^m,\mathbb{Z}_4)$ as the set of all m-variables quaternary functions.

Example 3.2. F: $\begin{array}{ccc}\mathbb{Z}_4^2 & \to & \mathbb{Z}_4 \\ (x_1,x_2) & \mapsto & x_1+x_2 \mod 4\end{array}$

The truth table is
$[F(0,0),F(0,1),F(0,2),F(0,3),F(1,0),F(2,0),F(3,0),F(1,1),F(1,2),F(1,3),$
$F(2,1),F(3,1),F(2,2),F(2,3),F(3,2),F(3,3)]$ i.e
$[0,1,2,3,1,2,3,2,3,0,3,0,0,1,1,2]$.

The support of F is defined by $supp(F)=\{u\in\mathbb{Z}_4^m \mid F(u)\neq 0\}$. We define the relative support of F by $supp_j(F)=\{u\in\mathbb{Z}_4^m \mid F(u)=j\}$ for all j in $\mathbb{Z}_4$ and $\eta_j(F)$ its size. The HAMMING weight $w_H(F)$ of F is the size of its support and the HAMMING distance between two m-variables quaternary functions F and G is $d_H(F,G)=w_H(F-G)$. The LEE weight $w_L(F)$ of F is $\eta_1(F)+\eta_3(F)+2\eta_2(F)$ and the LEE distance between two m-variables quaternary functions F and G is $d_L(F,G)=w_L(F-G)$. The Walsh transform of an m-variables quaternary function F is the complex mapping from $\mathbb{Z}_4^m$ to $\mathbb{C}$ defined by $W_F(u)=\sum_{v\in\mathbb{Z}_4^m} i^{u\cdot v+F(v)}$ where $u\cdot v$ denotes the usual inner product in $\mathbb{Z}_4^m \pmod 4$. We define $W_F^2(u)=\sum_{v\in\mathbb{Z}_4^m} i^{u\cdot v}(-1)^{F(v)}$ and $W_F^3(u)=\sum_{v\in\mathbb{Z}_4^m} i^{u\cdot v}(-i)^{F(v)}$

3.2. Quaternary Balancedness and Nonlinearity

Definition 3.3 (Balancedness)**.** Let $F\in\mathcal{F}(\mathbb{Z}_4^m,\mathbb{Z}_4)$.

$$F \text{ is balanced } \iff \forall j\in\mathbb{Z}_4,\ \eta_j(F)=4^{m-1}.$$

Let us give a balancedness characterisation of quaternary function.

Proposition 3.4. *Let $F\in\mathcal{F}(\mathbb{Z}_4^m,\mathbb{Z}_4)$.*

$$F \textit{ is balanced } \iff W_F(0)=W_F^2(0)=0.$$

Proof. By definition we have $W_F(0)=\sum_{v\in\mathbb{Z}_4^m} i^{F(v)}$ and $W_F^2(0)=\sum_{v\in\mathbb{Z}_4^m}(-1)^{F(v)}$, then $W_F(0)=\eta_0(F)-\eta_2(F)+i(\eta_1(F)-\eta_3(f))$ and $W_F^2(0)=\eta_0(F)-\eta_1(F)+\eta_2(F)-\eta_3(F)$. These two equalities give us 3 equations on η_j $(0\le j\le 3)$ by extracting real and imaginary parts. Since $\sum_{j\in\mathbb{Z}_4}\eta_j(F)=4^m$ we then obtain a system of 4 simultaneous equations in 4 unknowns that we solve. This finishes the proof. □

Similary to the binary case, we define the nonlinearity of quaternary function.

Definition 3.5 (Nonlinearity)**.** Let $F\in\mathcal{F}(\mathbb{Z}_4^m,\mathbb{Z}_4)$. The nonlinearity of F is defined by the minimum distance to all affine functions with
$nl_4^H(F)=\min_{G\ affine} d_H(F,G)$ under the HAMMING metric
and with $nl_4^L(F)=\min_{G\ affine} d_L(F,G)$ under the LEE metric.

Go on with a nonlinearity characterisation of quaternary function.

Proposition 3.6. *Let $F \in \mathcal{F}(\mathbb{Z}_4^m, \mathbb{Z}_4)$. The nonlinearity of F under the* HAMMING *metric is completely characterised by*

$$\begin{aligned} nl_4^H(F) &= 3 \cdot 4^{m-1} - \frac{1}{4} \max_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \left\{ 2Re(i^b W_F(a)) + (-1)^b W_F^2(2a) \right\} \\ &= 3 \cdot 4^{m-1} - \frac{1}{4} \max_{a \in \mathbb{Z}_4^m} \left\{ 2 \mid Re(W_F(a)) \mid + W_F^2(2a), 2 \mid Im(W_F(a)) \mid - W_F^2(2a) \right\} \end{aligned}$$

where $Re(z)$ and $Im(z)$ denote respectively the real and imaginary part of the complex number z.

Proof. By Definition 3.5, we have

$$nl_4^L(F) = \min_{G \ affine} d_H(F, G) = \min_{G \ affine} w_H(F - G)$$

Let S be the function in $\mathcal{F}(\mathbb{Z}_4^m, \mathbb{Z}_4)$ such that $S(u) = F(u) + a \cdot u + b$ with a in $\mathbb{Z}_4^m$ and b in $\mathbb{Z}_4$.

$$\begin{aligned} nl_4^H(F) &= \min_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \{\eta_1(S) + \eta_2(S) + \eta_3(S)\} \\ &= 4^m - \max_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \eta_0(S). \end{aligned}$$

Using the decomposition of $W_S(0)$, $W_S^2(0)$, $W_S^3(0)$ and the fact that $\eta_0(S)+\eta_1(S)+\eta_2(S)+\eta_3(S) = 4^m$ we obtain

$$\begin{aligned} \eta_0(S) &= \frac{1}{4} \left[4^m + W_S(0) + W_S^2(0) + W_S^3(0) \right] \\ &= \frac{1}{4} \left[4^m + \sum_{u \in \mathbb{Z}_4^m} \left(i^{F(u)+a \cdot u+b} + (-1)^{F(u)+a \cdot u+b} + (-i)^{F(u)+a \cdot u+b} \right) \right] \\ &= \frac{1}{4} \left[4^m + i^b W_F(a) + (-1)^b W_F^2(2a) + \overline{i^b W_F(a)} \right] \\ &= \frac{1}{4} \left[4^m + 2Re(i^b W_F(a)) + (-1)^b W_F^2(2a) \right]. \end{aligned}$$

The proof is completed, the second expression of $nl_4^H(F)$ is obvious using properties of complex numbers. □

Proposition 3.7. *Let $F \in \mathcal{F}(\mathbb{Z}_4^m, \mathbb{Z}_4)$. The nonlinearity of F under the* LEE *metric is completely characterised by*

$$\begin{aligned} nl_4^L(F) &= 4^m - \max_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \left\{ Re(i^b W_F(a)) \right\} \\ &= 4^m - \max_{a \in \mathbb{Z}_4^m} \{ \mid Re(W_F(a)) \mid, \mid Im(W_F(a)) \mid \}. \end{aligned}$$

Proof. By Definition 3.5, we have

$$nl_4^L(F) = \min_{G \ affine} d_L(F, G) = \min_{G \ affine} w_L(F - G).$$

Let S be the function in $\mathcal{F}(\mathbb{Z}_4^m, \mathbb{Z}_4)$ such that $S(u) = F(u) + a \cdot u + b$ with a in $\mathbb{Z}_4^m$ and b in $\mathbb{Z}_4$.

$$nl_4^L(F) = \min_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \{\eta_1(S) + 2\eta_2(S) + \eta_3(S)\}.$$

Using the decomposition of $W_S(0)$ and $W_S^3(0)$ we have

$$W_S(0) + W_S^3(0) = 2(\eta_0(S) - \eta_2(S)).$$

Moreover

$$W_S(0) + W_S^3(0) = 2Re(i^b W_F(a)).$$

As $\sum_{j \in \mathbb{Z}_4} \eta_j(S) = 4^m$, we obtain

$$\eta_1(S) + 2\eta_2(S) + \eta_3(S) = 4^m + \eta_2(S) - \eta_0(S).$$

That is

$$\begin{aligned} nl_4^L(F) &= \min_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \{4^m + \eta_2(S) - \eta_0(S)\} \\ &= 4^m - \max_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \left\{Re(i^b W_F(a))\right\} \end{aligned}$$

which ends the proof of the first expression. The second expression of $nl_4^L(F)$ is obvious by properties of complex numbers. □

3.3. Quaternary Bent Functions Properties

Definition 3.8 (Quaternary Bent functions)**.** Let F be an m-variables quaternary function. F is Bent if and only if $|W_F(a)| = 2^m$, for any $a \in \mathbb{Z}_4^m$.

Remark 3.9. If F is Bent, we have $W_F(a) = \pm 2^m$ or $W_F(a) = \pm i 2^m$.

Proof. $\widehat{\chi_F}(a) = \mu + i\nu$ with (μ, ν) in $\mathbb{Z} \times \mathbb{Z}$ and $|\widehat{\chi_F}(a)| = 2^m$.
Then $(2^m)^2 = \mu^2 + \nu^2$ and the only possible values for μ and ν are $(\mu, \nu) = (\pm 2^m, 0)$ or $(\mu, \nu) = (0, \pm 2^m)$: proved by recurrence on m for (μ, ν) in $\mathbb{N} \times \mathbb{N}$. □

Let us now focus on the maximal nonlinearity of an m-variables quaternary Bent function F according to the Hamming metric and the Lee metric respectively as shown in Fig.1.

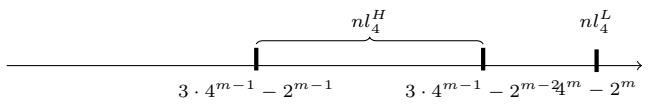


Figure 1. Nonlinearity of Quaternary Bent function .

Theorem 3.10. *Let F be a m-variables Bent function.*
(1) $3 \cdot 4^{m-1} - 2^{m-1} \leq nl_4^H(F) \leq 3 \cdot 4^{m-1} - 2^{m-2}$.
(2) $nl_4^H(F) = 3 \cdot 4^{m-1} - 2^{m-2}$ *if and only if* $W_F^2(2a) = \pm 2^m$.

Proof. (1): Proposition 3.6 gives $nl_4^H(F)$ equal to
$3 \cdot 4^{m-1} - \frac{1}{4} \sup_{a \in \mathbb{Z}_4^m} \{2 \mid Re(W_F(a)) \mid + W_F^2(2a), 2 \mid Im(W_F(a)) \mid - W_F^2(2a)\}$.
Let us write $nl_4^H(F) = 3 \cdot 4^{m-1} - \frac{1}{4}y$
where $y = \sup_{a \in \mathbb{Z}_4^m} \{2 \mid Re(W_F(a)) \mid + W_F^2(2a), 2 \mid Im(W_F(a)) \mid - W_F^2(2a)\}$
and $x = W_F^2(2a) = \sum_{u \in \mathbb{Z}_4^m} (i)^{2a.u}(-1)^{F(u)} = \sum_{u \in \mathbb{Z}_4^m} (-1)^{a.u}(-1)^{F(u)}$.
As F is Bent, we use Remark 3.9 to distinguish two main cases in order to evaluate y (let c=2^{m+1}):

- $W_F(a) = \pm 2^m : y = Max\{c + x, -x\}$
- $W_F(a) = \pm i2^m : y = Max\{x, c - x\}$

The geometric representation of y in terms of x (Fig.2) shows that y ranges between 2^m and 2^{m+1} which completes the proof.

(2): Let $nl_4^H(F) = 3 \cdot 4^{m-1} - 2^{m-2}$. If F is Bent and $W_F(a)$ real then $2^m = \sup_{a \in \mathbb{Z}_4^m} \{2^{m+1} + W_F^2(2a), -W_F^2(2a)\}$. In this case $W_F^2(2a) < 0$ and $W_F^2(2a)$ is equal to -2^m or $2^{m+1} + W_F^2(2a) = 2^m$ that is $W_F^2(2a) = 2^m - 2^{m+1} = -2^m$. The case $W_F(a)$ is imaginary is similar.

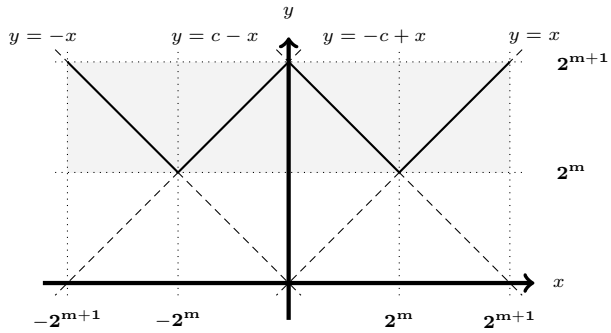


Figure 2. y in terms of x.

□

Theorem 3.11. *Let F be an m-variables Bent function.*

$$nl_4^L(F) = 4^m - 2^m.$$

Proof. Proposition 3.7 gives $nl_4^L(F) = 4^m - \max_{a \in \mathbb{Z}_4^m, b \in \mathbb{Z}_4} \{Re(i^b W_F(a))\}$.
Using remark 3.9 we have $Re(i^b W_F(a)) = \pm 2^m$ which finishes the proof. □

4. Galois Rings and Cyclotomic Classes

In this section we give definitions and properties of the Galois ring $GR(4, m)$ without proofs. We refer the reader to [20] and [8] for further informations about Galois rings.

4.1. Galois Rings

As usual, $\mathbb{Z}_4 = \mathbb{Z}/4\mathbb{Z} = \{0, 1, 2, 3\}$ is the ring of integers modulo 4 and $\mathbb{F}_2$ the finite field with two elements. Let $\mu : \mathbb{Z}_4 \to \mathbb{F}_2$ be the mod-2 reduction map. We extend μ to $\mathbb{Z}_4[x] \to \mathbb{F}_2[x]$ in the natural way. A monic polynomial $h(x)$ in $\mathbb{Z}_4[x]$ of degree m is said to be basic irreducible if $h_2(x) = \mu(h(x))$ is a monic irreducible primitive divisor of $x^{2^m-1} - 1$ in $\mathbb{F}_2[x]$ (Hensel lift). The Galois ring $R = GR(4, m)$ of 4^m elements is a Galois extension of order m of $\mathbb{Z}_4$ and is isomorphic to the factor ring $\mathbb{Z}_4[x]/(h(x))$ where $h(x)$ is a monic basic irreducible polynomial of degree m (b-polynomial). Let β be a root of $h(x)$ of order $2^m - 1$ ($\beta^{2^m-1} - 1 = 0$). Then R is the polynomial ring $\mathbb{Z}_4[\beta]$ where $\{1, \beta, \cdots, \beta^{m-1}\}$ is a basis of R over $\mathbb{Z}_4$. The Galois ring R is a local ring having a unique maximal ideal $D = 2R$ made up of the 2^m zero divisors. The residue class field $K = R/D$ is isomorphic to the finite field $\mathbb{F}_{2^m}$ under the canonical map $z \mapsto \bar{z}$ from R to K. The Teichmüller system $\mathcal{T} = \{0, 1, \beta, \cdots, \beta^{2^m-2}\}$ is the set of roots of $x^{2^m} - x$ in R and can be viewed as the set of representatives of K as $D = 2R = 2\mathcal{T}$. Let $\theta = \bar{\beta}$ be a primitive root of $h_2(x)$ in $\mathbb{F}_2[x]$, we can identify K with $\mathbb{F}_{2^m} = \overline{\mathcal{T}} = \{0, 1, \theta, \cdots, \theta^{2^m-2}\}$. The multiplicative group $R^\star = R \setminus D$ of R is a group of order $(2^m - 1)2^m$ which is the direct product $\mathcal{H} \times \mathcal{U}$ where $\mathcal{H}$ is the cyclic group of order $(2^m - 1)$ generated by β and $\mathcal{U}$ is the Abelian group of principal units of R of order 2^m that is elements of the form $1 + 2z_0$ with z_0 in $\mathcal{T}$. There are two canonical ways to represent the 4^m elements of R, a multiplicative one and an additive one. In the multiplicative representation, every element z of R has a unique expansion $z = z_1 + 2z_2$ with z_1 and z_2 in $\mathcal{T}$.

4.2. Cyclotomic Classes

Let $R = GR(4, m)$ be the Galois ring of 4^m elements, $D = \{0, 2, 2\beta, \cdots, 2\beta^{2^m-2}\}$ the set of zero divisors with $\mid D \mid = 2^m$ and $R^\star = \{z_1(1+2z_0),\ z_0 \in \mathcal{T}, z_1 \in \mathcal{T} \setminus \{0\}\}$ the multiplicative group of R with $\mid R^\star \mid = 2^m(2^m - 1)$.

Definition 4.1. Let m be a natural integer and $R = GR(4, m)$ be the Galois ring of 4^m elements and $R^\star$ its multiplicative group. The 2^m cyclotomic classes of order $2^m - 1$ of $R^\star$ are:
$C_k = \{\beta^j + 2\beta^k,\ 0 \le j \le 2^m - 2\}$ for any k such that $0 \le k \le 2^m - 2$ and $C_{2^m-1} = \{\beta^j,\ 0 \le j \le 2^m - 2\}$.

5. Construction

Let $R = GR(4, m)$ be the Galois ring of 4^m elements and D the set of zero divisors of R. Let us consider the 2^m cyclotomic classes C_k of order $2^m - 1$ of $R^\star$ (see Definition 4.1).
We construct the quaternary function F such that the function F takes the same value for each element of C_k.

We now compute formally the expressions of W_F and W_F^2 for this constructed function F. We have

$$\begin{cases} C_k & = & \{\beta^j + 2\beta^k,\ 0 \leq j \leq 2^m - 2\},\ 0 \leq k \leq 2^m - 2 \\ \\ C_{2^m-1} & = & \{\beta^j,\ 0 \leq j \leq 2^m - 2\} \end{cases}$$

with $\mid C_{k,\ 0\leq k\leq 2^m-1} \mid = 2^m - 1$. And $D = \{0\} \cup \{2\beta^j,\ 0 \leq j \leq 2^m - 2\}$.
Let a in $\mathbb{Z}_4^m$.

$$W_F(a) = \underbrace{\sum_{v\in D} i^{a\cdot v+F(v)}}_{S_D(a)} + \sum_{0\leq k\leq 2^m-2} \underbrace{\left(\sum_{v\in C_k} i^{a\cdot v+F(v)}\right)}_{S_{C_k}(a)} + \underbrace{\sum_{v\in C_{2^m-1}} i^{a\cdot v+F(v)}}_{S_{C_{2^m-1}}(a)}$$

$$W_F(a) = S_D(a) + \sum_{0\leq k\leq 2^m-2} S_{C_k}(a) + S_{C_{2^m-1}}(a) \tag{5.1}$$

As $D = \{0, 2, 2\beta, \cdots, 2\beta^{2^m-2}\}$ we have

$$S_D(a) = i^{F(0)} + \sum_{0\leq k\leq 2^m-2} (-1)^{a\cdot\beta^k} i^{F(2\beta^k)} \tag{5.2}$$

If $v \in C_k$ for $0 \leq k \leq 2^m - 2$ then $v = \beta^j + 2\beta^k$ with $0 \leq j \leq 2^m - 2$ and

$$S_{C_k}(a) = (-1)^{a\cdot\beta^k} \sum_{0\leq j\leq 2^m-2} i^{a\cdot\beta^j} i^{F(\beta^j+2\beta^k)} \tag{5.3}$$

If $v \in C_{2^m-1}$ then $v = \beta^j$ for $0 \leq j \leq 2^m - 2$ and

$$S_{C_{2^m-1}}(a) = \sum_{0\leq j\leq 2^m-2} i^{a\cdot\beta^j} i^{F(\beta^j)} \tag{5.4}$$

In equations (5.2), (5.3) and (5.4), terms of the form $(-1)^{a\cdot\beta^k}$ and $i^{a\cdot\beta^j}$ show that $W_F(a)$ depends on the b-polynomial used to construct the Galois ring and terms of the form $i^{F(2\beta^k)}$, $i^{F(\beta^j+2\beta^k)}$ and $i^{F(\beta^j)}$ show that $W_F(a)$ depends on the way that F takes value on the different cosets $C_{k,\ 0\leq k\leq 2^m-1}$ and D.
As the construction states that for a given class C_k, the function F takes the same value, if $v = \beta^j + 2\beta^k \in C_k$ then let us define $F_k = F(v) = F(\beta^j + 2\beta^k)$ which does not depend on j.

$$\begin{aligned} W_F(a) &= S_D(a) + \sum_{0\leq k\leq 2^m-2} \left((-1)^{a\cdot\beta^k} i^{F_k} \sum_{0\leq j\leq 2^m-2} i^{a\cdot\beta^j} \right) + i^{F_{2^m-1}} \sum_{0\leq j\leq 2^m-2} i^{a\cdot\beta^j} \\ &= S_D(a) + \left(\sum_{0\leq j\leq 2^m-2} i^{a\cdot\beta^j} \right) \left(\sum_{0\leq k\leq 2^m-2} (-1)^{a\cdot\beta^k} i^{F_k} + i^{F_{2^m-1}} \right) \end{aligned}$$

We have now to distinguish the case $a \in D$ from the case $a \notin D$.
$\underline{a \in D}$
Let $a = 2\beta^l$ with $0 \leq l \leq 2^m - 2$ or $a = 0$.

$$W_F(0) = \sum_{v\in D} i^{F(v)} + (2^m - 1) \sum_{0\leq k\leq 2^m-1} i^{F_k} \tag{5.5}$$

$$W_F(a) = \sum_{v\in D} i^{F(v)} + \left(\sum_{0\le j\le 2^m-2} (-1)^{\beta^l\cdot\beta^j}\right)\left(\sum_{0\le k\le 2^m-1} i^{F_k}\right) \tag{5.6}$$

$\underline{a \notin D}$

Let $a = \beta^s + 2\beta^l$ in C_l for $0 \le l \le 2^m - 2$ and for $l = 2^m - 1$ we have $a = \beta^s$ with $0 \le s \le 2^m - 2$.

Furthermore : $(-1)^{a\cdot\beta^k} = (-1)^{\beta^s\cdot\beta^k}$ and $i^{a\cdot\beta^j} = \begin{cases} i^{\beta^s\cdot\beta^j}(-1)^{\beta^l\cdot\beta^j} \\ i^{\beta^s\cdot\beta^j} \end{cases}$

That is

$$W_F(a) = S_D(s) + A(s,l)\left(B(s) + i^{F_{2^m-1}}\right) \tag{5.7}$$

where

$$S_D(s) = i^{F(0)} + \sum_{0\le k\le 2^m-2} (-1)^{\beta^s\cdot\beta^k} i^{F(2\beta^k)} \tag{5.8}$$

$$A(s,l) = \sum_{0\le j\le 2^m-2} (-1)^{\beta^l\cdot\beta^j} i^{\beta^s\cdot\beta^j} \tag{5.9}$$

$$B(s) = \sum_{0\le k\le 2^m-2} (-1)^{\beta^s\cdot\beta^k} i^{F_k} \tag{5.10}$$

$$A(s) = \sum_{0\le j\le 2^m-2} i^{\beta^s\cdot\beta^j} \tag{5.11}$$

We have that $S_D(s)$ and $B(s)$ do not depend on the class of a but only on values of F and A depends only on a but not on values of F. Moreover, we have

$$W_F^2(2a) = \sum_{v\in\mathbb{Z}_4^m} (-1)^{a.v}(-1)^{F(v)} = \sum_{v\in D}(-1)^{F(v)} + \sum_{0\le k\le 2^m-1}\left(\sum_{v\in C_k}(-1)^{a\cdot v+F(v)}\right)$$

As for the calculation of $W_F(a)$, we find that

$$W_F^2(2a) = \sum_{v\in D}(-1)^{F(v)} + \sum_{0\le k\le 2^m-1}(-1)^{F_k}\sum_{0\le j\le 2^m-2}(-1)^{a.\beta^j} \tag{5.12}$$

Equations (5.6) and (5.7) give the exact value of $W_F(a)$ and (5.12) the exact value of $W_F^2(2a)$ according to the detailed calculation done by equations (5.1)–(5.5) and (5.8)–(5.11).

Remark 5.1 (Balancedness of F)**.** The balancedness of F depends on the way that F takes value on the different cosets C_k and D.

Remark 5.2 (Nonlinearity of F)**.** The nonlinearity of F under the Hamming and Lee metric depends on the choice of the b-polynomial and the way that F takes value according to u belongs to C_k or D.

We now apply these results to a particular configuration in order to obtain a balanced quaternary function with high nonlinearity under the Hamming metric and Lee metric as shown in Fig.3 and Fig.4 respectively.

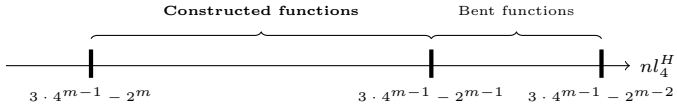


Figure 3. Constructed Quaternary function nonlinearity nl_4^H.

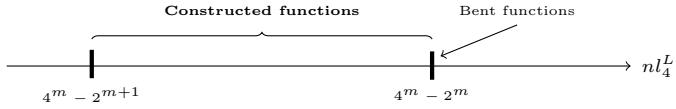


Figure 4. Constructed Quaternary function nonlinearity nl_4^L.

Proposition 5.3. *For k, $0 \leq k \leq 2^m - 2$ and for $\gamma \in \{0, 1, 2, 3\}$, we define $\delta_k = \gamma$ if $k \equiv \gamma \pmod 4$. With a suitable b-polynomial, we construct an m-variables quaternary function F as follows: $F(\beta^j + 2\beta^k) = F(2\beta^k) = \delta_k$ and $F(0) = 3$, for j, $0 \leq j \leq 2^m - 2$. This quaternary function is balanced and its nonlinearity under the* Hamming *metric satisfies $3 \cdot 4^{m-1} - 2^m \leq nl_4^H(F) \leq 3 \cdot 4^{m-1} - 2^{m-1}$ and its nonlinearity under the* Lee *metric satisfies $4^m - 2^{m+1} \leq nl_4^L(F) \leq 4^m - 2^m$.*

Numerical Results

Nonlinearity under the Hamming metric $nl_4^H(F)$ and the Lee metric $nl_4^L(F)$ (using Propositions 3.6 and 3.7) of constructed balanced quaternary m-variables functions F with Nbp the number of possible b-polynomials, Nbs the number of suitable b-polynomials and $B_1^H = 3 \cdot 4^{m-1} - 2^m, B_2^H = 3 \cdot 4^{m-1} - 2^{m-1}, B_1^L = 4^m - 2^{m+1}$ and $B_2^L = 4^m - 2^m$.

m	Nbp	Nbs	*suitable bpolynomial*	B_1^H	$\mathbf{nl_4^H(F)}$	B_2^H	B_1^L	$\mathbf{nl_4^L(F)}$	B_2^L
3	2	2	$x^3 + 2x^2 + x + 3$	40	**44**	44	48	**56**	56
4	2	2	$x^4 + 2x^2 + 3x + 1$	176	**180**	184	224	**232**	240
5	6	6	$x^5 + 3x^2 + 2x + 3$	736	**744**	752	960	**976**	992
6	6	2	$x^6 + x^5 + x^4 + 2x^2 + 3x + 1$	3008	**3032**	3040	3968	**4016**	4032
7	18	14	$x^7 + 2x^4 + x + 3$	12160	**12208**	12224	16128	**16224**	16256
8	16	2	$x^8 + 3x^5 + x^3 + 2x^2 + 3x + 1$	48896	**49008**	49024	65024	**65248**	65280
9	48	10	$x^9 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3$	196096	**196288**	196352	261120	**261504**	261632

6. Derived Boolean Functions

Let $R = GR(4, m)$ be the Galois ring of 4^m elements and D the set of zero divisors of R. Let us consider the m-variables quaternary function F obtained by the construction which uses the 2^m cyclotomic classes of order $2^m - 1$ of $R^\star$. By taking the binary images of F under the Gray map, we obtain $n = 2m$-variables Boolean functions which are balanced and having high nonlinearity.

Definition 6.1. The Gray map ϕ is defined from $\mathbb{Z}_4$ to $\mathbb{F}_2 \times \mathbb{F}_2$ with $\phi(2q + r) = (q, q \oplus r)$. We also define Q from $\mathbb{Z}_4$ to $\mathbb{F}_2$ with $Q(2q + r) = q$.

The Gray map is clearly a bijection from $\mathbb{Z}_4$ to $\mathbb{F}_2^2$ and its inverse is defined by $\phi^{-1}(q, s) = 2q + (q \oplus s)$. Identifying $\mathbb{F}_2^m \times \mathbb{F}_2^m$ to $\mathbb{F}_2^{2m}$, we extend naturally ϕ

to $\mathbb{Z}_4^m$ componentwise by $\phi_m(2q_0 + r_0, \cdots, 2q_{m-1} + r_{m-1}) = (q_0, \cdots, q_{n-1}, q_0 \oplus r_0, \cdots, q_{m-1} \oplus r_{m-1})$ and ϕ_m^{-1} to $\mathbb{F}_2^{2m}$ by $\phi_m^{-1}(q_0, \cdots, q_{n-1}, s_0, \cdots, s_{m-1}) = (2q_0 + q_0 \oplus s_0, \cdots, 2q_{m-1} + q_{m-1} \oplus s_{m-1})$.

Definition 6.2. The $2m$-variables Boolean function f derived by the Gray map is

$$\begin{array}{rccl} f: & \mathbb{F}_2^{2m} & \to & \mathbb{F}_2 \\ & y & \mapsto & Q(F(\phi_m^{-1}(y))) \end{array}$$

Numerical Results :

Nonlinearity $nl_2(f) = 2^{n-1} - \frac{1}{2} \max_{a \in \mathbb{F}_2^n} |\boldsymbol{W}_f(a)|$ of obtained balanced n-variables Boolean functions f with $n = 2m$ derived from the previous constructed balanced quaternary m-variables functions F compared with known values.

n	bnl^1	bnl^2	nl^1	nl^2	$\mathbf{nl_2(f)}$
6	12	10	22	24	**24**
8	58	70	94	112	**112**
10	260	366	390	478	**464**
12	1124	1700	1600		**1952**
14	4760	7382	6524		**8000**

bnl^1: Lobanov's lower bound
bnl^2: Carlet - Feng's ([4]) lower bound

nl^1: Best balanced exact Nonlinearity before
nl^2: Carlet-Feng's ([4]) exact Nonlinearity with optimal algebraic immunity

7. Numerical Example for $m = 3$ and $n = 6$

Let consider the Galois ring $R = GR(4, 3)$ of 64 elements built with the b-polynomial $h(x) = x^3 + 2x^2 + x + 3$.

The 8 cyclotomic classes of order 7 of $R^\star$ (Def.4.1) are :

C_0 : $\{3, \beta + 2, \beta^2 + 2, \beta^3 + 2, \beta^4 + 2, \beta^5 + 2, \beta^6 + 2\}$
C_1 : $\{1 + 2\beta, 3\beta, \beta^2 + 2\beta, \beta^3 + 2\beta, \beta^4 + 2\beta, \beta^5 + 2\beta, \beta^6 + 2\beta\}$
C_2 : $\{1 + 2\beta^2, \beta + 2\beta^2, \beta^2 + 2\beta^2, \beta^3 + 2\beta^2, \beta^4 + 2\beta^2, \beta^5 + 2\beta^2, \beta^6 + 2\beta^2\}$
C_3 : $\{1 + 2\beta^3, \beta + 2\beta^3, \beta^2 + 2\beta^3, \beta^3 + 2\beta^3, \beta^4 + 2\beta^3, \beta^5 + 2\beta^3, \beta^6 + 2\beta^3\}$
C_4 : $\{1 + 2\beta^4, \beta + 2\beta^4, \beta^2 + 2\beta^4, \beta^3 + 2\beta^4, \beta^4 + 2\beta^4, \beta^5 + 2\beta^4, \beta^6 + 2\beta^4\}$
C_5 : $\{1 + 2\beta^5, \beta + 2\beta^5, \beta^2 + 2\beta^5, \beta^3 + 2\beta^5, \beta^4 + 2\beta^5, \beta^5 + 2\beta^5, \beta^6 + 2\beta^5\}$
C_6 : $\{1 + 2\beta^6, \beta + 2\beta^6, \beta^2 + 2\beta^6, \beta^3 + 2\beta^6, \beta^4 + 2\beta^6, \beta^5 + 2\beta^6, \beta^6 + 2\beta^6\}$
C_7 : $\{1, \beta, \beta^2, \beta^3, \beta^4, \beta^5, \beta^6\}$

The obtained partitions applying the construction are (Prop. 5.3) :
Let $E_j = \{u \in \mathbb{Z}_4^3, F(u) = j\}$, $j = 0, 1, 2, 3$.

$E_0 = C_0 \cup C_4 \cup \{2, 2\beta^4\}$
$E_1 = C_1 \cup C_5 \cup \{2\beta, 2\beta^5\}$
$E_2 = C_2 \cup C_6 \cup 2\{\beta^2, 2\beta^6\}$.
$E_3 = C_3 \cup C_7 \cup \{2\beta^3, 0\}$

The balanced constructed 3-variables quaternary function F is :
$F = 332231231100120031200311130220320022002133113213032113030122302$
with $nl_4^H(F) = 44$ and $nl_4^L(F) = 56$.

The balanced derived 6-variables Boolean function f is :
$f = 11111011000001001010010001011011001100101100110101100010100111101$
with $nl_2(f) = 24$.

8. Conclusion

This paper presents new results on quaternary cryptographic functions, bringing out a new approach of functions used in the security of pseudo-random generators of stream and block ciphers. The main goal of this work, similarly motivated by the $\mathbb{Z}_4$ linearity paper [8], is to present an alternative to the open problem of finding optimal Boolean functions. After defining quaternary functions and describing their $\mathbb{Z}_4$ balancedness and nonlinearity, under the Hamming metric and the Lee metric, we give results on the maximal nonlinearity of quaternary Bent functions. Using the algebraic structure of a Galois ring, we present a general construction of quaternary functions, pointing out necessary trade offs in order to obtain optimal cryptographic properties. In a natural way, we apply this construction with a particular configuration to get balanced and high nonlinearity quaternary functions. Faithful to our main objective, we take the image of our quaternary constructed functions under the Gray map to obtain balanced and high nonlinearity Boolean functions. $\mathbb{Z}_4$ codes, Galois Rings and Difference Sets over $\mathbb{Z}_4$ seems to offer great investment opportunities and reinforce our motivation to go on with this new kind of approach.

References

[1] A. Canteau, M. Trabbia: *Improved fast correlation attacks using parity-check equations of weight 4 and 5*, Advances in Cryptology, EUROCRYPT 2000, LNCS, **1807** (2000), 573–588.

[2] C. Carlet: *On Bent and highly nonlinear balanced/resilient functions and their algebraic immunities*, AAECC - LNCS, **3857** (2006), 1–28.

[3] C. Carlet, S. Dubuc: *On generalized Bent and q-ary perfect nonlinear functions*, Finite Fields and Applications **1999**, (2001), 81–94.

[4] C. Carlet, K. Feng: *An infinite class of balanced functions with optimal algbreaic immunity, good immunity to fast algebraic attacks and good nonlinearity*, Advances in Cryptology, ASIACRYPT 2008, LNCS, **5350** (2008), 425-440.

[5] N. Courtois, W. Meier: *Algebraic attacks on stream ciphers with linear feedback*, Advances in Cryptology, EUROCRYPT 2003, LNCS, **2656** (2003), 345–359.

[6] N. Courtois, J. Pieprzyk: *Cryptanalysis of block ciphers with over-defined systems of equations*, Advances in Cryptology, ASIACRYPT 2002, LNCS, **2501** (2002), 267–287.

[7] C. Ding, G. Xiao, W. Shan: *The Stability Theory of Stream Ciphers*, LNCS, Springer, Berlin, 1991.

[8] A. R. Hammons, P. V. Kumar, A. R. Calderbank, N. J .A. Sloane, P. Solé: *The $\mathbb{Z}_4$ linearity of Kerdock, Preparata, Goethals and related codes*, IEEE Transactions on Information Theory **40** (2) (1994), 301–319.

[9] X. D. Hou: *p-ary and q-ary versions of certain results about Bent functions and resilient functions*, Finite Fields and Applications **10** (2004), 566–582.

[10] X. D. Hou: *q-ary Bent functions constructed from chain rings*, Finite Fields and Applications **4** (1998), 55–61.

[11] X. D. Hou: *Bent functions, partial difference sets and quasi-Frobenius rings*, Designs, Codes and Cryptography **20** (2000), 251–268.

[12] P. V. Kumar, T. Hellesth , A. R. Calderbank, A. R. Hammons: *Large Families of Quaternary Sequences with Low Correlation*, IEEE Transactions on Information Theory **42** (2) (1996), 579–592.

[13] S. Kavut, S. Maitra, S. Sarkar, M. D. Yücel: *Enumeration of 9-variables rotation symetric Boolean functions having non-linearity* > 240, Advances in Cryptology, INDOCRYPT 2006, LNCS, **4329** (2006), 266–279.

[14] P. V. Kumar, R. A. Scholtz, L. R. Welch: *Generalized Bent Functions and Their Properties*, Journal of Combinatorial Theory, Ser. A, **1** (40) (1985), 90–107.
[15] N. Li, W. F. Qi: *Construction and analysis of Boolean functions of* $2t + 1$ *variables with maximum algebraic immunity*, Advances in Cryptology, ASIACRYPT 2006, LNCS, **4284** (2006), 84–98.
[16] W. Meier, E. Pasalic, C. Carlet: *Algebraic attacks and decomposition of Boolean functions*, Advances in Cryptology, EUROCRYPT 2004, LNCS, **3027** (2004), 474–491.
[17] O. S. Rothaus: *On Bent functions*, Journal of Combinatorial Theory **20** (1976), 300–305.
[18] Z. Saber, M. F. Uddin and A. Youssef: *On the existence of* $(9, 3, 5, 240)$ *resilient functions*, IEEE Transactions on Information Theory **52** (5) (2006), 2269–2270.
[19] P. Solé and N. Tokareva: *Connections between quaternary and binary Bent functions*, Cryptology ePrint Archives, *http://www.eprint.iacr.org/2009/544*, 2009.
[20] B. R. McDonald: *Finite Rings with Identity*, Marcel Dekker Inc., 1974.

Zoubida Jadda, CREC, UR - MACCLIA, Ecoles Militaires Saint-Cyr Coëtquidan, France,
e-mail: zoubida.jadda@st-cyr.terre-net.defense.gouv.fr

Patrice Parraud, CREC, UR - MACCLIA, Ecoles Militaires Saint-Cyr Coëtquidan, France,
e-mail: patrice.parraud@st-cyr.terre-net.defense.gouv.fr

ŠIFROVACÍ METODA ZALOŽENÁ NA FRAKTÁLNÍ KOMPRESI

TOMÁŠ GRÍSA

Abstrakt. Tento článek se zabývá teoretickými principy fraktální komprese a využitím modifikovaného algoritmu fraktální komprese pro symetrickou kryptografii. Kompaktní matematická teorie a vlastní algoritmus navržené symetrické šifry (Fractal Compression Based Cryptosystem - FCBC) jsou zde prezentovány. Tato práce také obsahuje příklad ilustrující praktické využití této šifry a její různé modifikace.

1. Úvod

V posledních letech se ukázalo, že teorii fraktálů lze využít v mnoha teoretických i praktických oborech. Příkladem může být fraktální komprese obrazu využívající *systém iterovaných funkcí (IFS)* a *segmentovaný systém iterovaných funkcí (PIFS)*, kterou v roce 1987 navrhl Barnsley (viz. [1], [2]). Využil soběpodobnost pro obrazovou kompresi tak, že atraktor PIFS je aproximací původního obrazu. Avšak hledání soběpodobných částí v rastrovém obraze je výpočetně náročné.

Další zajímavou oblastí, kde lze aplikovat teorii fraktálů, je kryptografie. Tento text se zabývá právě touto oblastí. Popisuje modifikovaný reverzní algoritmus fraktální komprese použitý pro symetrickou šifru (FCBC), která je hlavním výsledkem této práce. V navržené šifře se PIFS iterativně aplikuje na libovolnou zprávu vhodné délky. Atraktorem tohoto PIFS je pak výsledná zašifrovaná zpráva. Klíč a zpráva k zašifrování jsou použity jako koeficienty daného PIFS. Zašifrovaná zpráva je tedy v jistém smyslu soběpodobná. Pokud známe klíč, můžeme pomocí jednoduchých výpočtů obdržet původní zprávu.

V textu se předpokládá znalost pojmů *metrický prostor, posloupnost, konvergence posloupnosti, cauchyovská posloupnost, úplný metrický prostor* a *kompaktní množina*.

2. Matematické principy

Definice 2.1. Nechť $(\mathcal{M}, d)$ je metrický prostor a $T : \mathcal{M} \to \mathcal{M}$. Zobrazení T nazveme:

- *lipschitzovské* právě tehdy, když existuje konstanta $l > 0$ taková, že:

$$d\left(T\left(x\right), T\left(y\right)\right) \leq l \cdot d\left(x, y\right), \qquad \forall x, y \in \mathcal{M},$$

2010 *MSC.* Primární 94A60; Sekundární 11T71.

Klíčová slova. Fraktální komprese, symetrický kryptosystém, Fractal Compression Based Cryptosystem, FCBC.

Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

- *kontraktivní* právě tehdy, když existuje konstanta $0 < c < 1$ taková, že:

$$d\left(T\left(x\right), T\left(y\right)\right) \leq c \cdot d\left(x, y\right), \qquad \forall x, y \in \mathcal{M},$$

- *neexpanzivní* právě tehdy, když:

$$d\left(T\left(x\right), T\left(y\right)\right) \leq d\left(x, y\right), \qquad \forall x, y \in \mathcal{M}.$$

Tvrzení 2.2. *Jestliže je zobrazení $T : \mathcal{M} \to \mathcal{M}$ lipschitzovské, potom je i spojité.*

Důkaz. Předpokládejme T spojité s lipschitzovým koeficientem l. Nechť $\varepsilon > 0$. Pokud $|x - y| < \varepsilon/l$, dostáváme že

$$|T(x) - T(y)| \leq l|x - y| < l \cdot \frac{\varepsilon}{l} = \varepsilon,$$

kde první nerovnost je ekvivalentní s lipschitzovou podmínkou. Pokud je tedy ε libovolné, pak T je spojité. □

Definice 2.3. Bod $x^* \in \mathcal{M}$ se nazývá *pevným bodem zobrazení* $T : \mathcal{M} \to \mathcal{M}$ právě tehdy, když

$$T\left(x^*\right) = x^* = \lim_{n \to \infty} T^n\left(x^*\right).$$

Věta 2.4. (Banachova věta o pevném bodu kontraktivního zobrazení) *Buď $\mathcal{M}$ úplný metrický prostor a $T : \mathcal{M} \to \mathcal{M}$ kontraktivní zobrazení. Potom T má právě jeden pevný bod.*

Důkaz. Zvolme libovolné $x \in \mathcal{M}$. Potom pro $n > m$ platí

$$d\left(T^m\left(x\right), T^n\left(x\right)\right) < c \cdot d\left(T^{m-1}\left(x\right), T^{n-1}\left(x\right)\right) < c^m \cdot d\left(x, T^{n-m}\left(x\right)\right). \tag{2.1}$$

Dále využijme trojúhelníkovou nerovnost:

$$\begin{aligned} d\left(x, T^k\left(x\right)\right) &\leq d\left(x, T^{k-1}\left(x\right)\right) + d\left(T^{k-1}\left(x\right), T^k\left(x\right)\right) \\ &\leq d\left(x, T\left(x\right)\right) + d\left(T\left(x\right), T\left(T\left(x\right)\right)\right) + \ldots + d\left(T^{k-1}\left(x\right), T^k\left(x\right)\right) \\ &\leq \left(1 + c + \ldots + c^{k-2} + c^{k-1}\right) \cdot d\left(x, T\left(x\right)\right) \\ &\leq \frac{1}{1-c} \cdot d\left(x, T\left(x\right)\right). \end{aligned} \tag{2.2}$$

Nyní můžeme přepsat výraz 2.1 následovně:

$$d\left(T^m\left(x\right), T^n\left(x\right)\right) < \frac{c^m}{1-c} \cdot d\left(x, T\left(x\right)\right).$$

Jestliže $c < 1$, potom můžeme vzít levou stranu výrazu libovolně malou, pokud jsou n a m dostatečně velká. To znamená, že posloupnost $x, T\left(x\right), T\left(T\left(x\right)\right), \ldots$ je cauchyovská. Pokud je $\mathcal{M}$ úplný metrický prostor, potom bod $x^* = \lim_{n\to\infty} T^n\left(x\right)$ leží v $\mathcal{M}$. Nyní podle tvrzení 2.2 platí, že pokud je T kontraktivní, je T zároveň spojité. Tudíž $T\left(x^*\right) = T\left(\lim T^n\left(x\right)\right) = \lim T^{n+1}\left(x\right) = x^*$.
Jednoznačnost dokažme následovně: předpokládejme, že x_1^* a x_2^* jsou dva pevné body. Potom $d\left(T\left(x_1^*\right), T\left(x_2^*\right)\right) = d\left(x_1^*, x_2^*\right)$, zároveň ale platí $d\left(T\left(x_1^*\right), T\left(x_2^*\right)\right) \leq c \cdot d\left(x_1^*, x_2^*\right)$. Tím dostáváme spor. □

Věta 2.5. (Kolážová věta) *Buď T kontraktivní zobrazení s koeficientem kontrakce c, a x^* pevným bodem zobrazení T. Potom platí:*

$$d(x, x^*) \leq \frac{1}{1-c} \cdot d(x, T(x))$$

Důkaz. Věta plyne z výrazu 2.2, pokud k jde k nekonečnu. □

Definice 2.6. Buď T lipschitzovské zobrazení. Pokud existuje $n \in \mathbb{N}$ takové, že T^n je kontraktivní, pak zobrazení T nazveme *podmíněně kontraktivní*. Exponent n nazveme *exponentem podmíněné kontrakce.*

Věta 2.7. (Zobecněná kolážová věta) *Buď T podmíněně kontraktivní s exponentem podmíněné kontrakce n, potom toto zobrazení má právě jeden pevný bod x^*, kterého dosáhneme pro libovolné x následovně:*

$$x^* = T(x^*) = \lim_{k \to \infty} T^k(x).$$

Potom platí:

$$d(x, x^*) \leq \frac{1}{1-c} \frac{1-l^n}{1-l} \cdot d(x, T(x)),$$

kde c je koeficient kontrakce T^n, a l je lipschitzovým koeficientem T.

Důkaz. Viz. [2], str. 36 □

Definice 2.8. Pokud je $(\mathcal{M}, d)$ úplný metrický prostor, potom *systémem iterovaných funkcí (IFS)* nazýváme konečnou množinu kontraktivních zobrazení $F = \{w_1, w_2, ..., w_n\}$ definovaných na $\mathcal{M}$.

Definice 2.9. Pro metrický prostor $(\mathcal{M}, d)$ označme $H(\mathcal{M})$ systém všech neprázdných kompaktních podmnožin $\mathcal{M}$. Potom zobrazení $d_h : H(\mathcal{M}) \times H(\mathcal{M}) \to \mathbb{R}_0^+$ definované předpisem

$$d_h(A, B) = \max\{\sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b)\}$$

pro neprázdné $A, B \subseteq \mathcal{M}$ nazveme *Hausdorffovou vzdáleností.* Ta splňuje axiomy metriky, tudíž $(H(\mathcal{M}), d_h)$ tvoří metrický prostor.

Věta 2.10. (IFS věta) *Je-li $(\mathcal{M}, F)$ systémem iterovaných funkcí, potom transformace $\mathcal{W} : H(\mathcal{M}) \to H(\mathcal{M})$, pro kterou platí*

$$\mathcal{W}(B) = \bigcup_{i=1}^{n} w_i(B)$$

pro všechna $B \in H(\mathcal{M})$, je kontraktivním zobrazením na $(H(\mathcal{M}), d_h)$ s koeficientem kontrakce $c = \max\{c_1, ..., c_n\}$. Pak tato transformace má jediný pevný bod $A \in H(\mathcal{M})$, který vyhovuje rovnici $A = \mathcal{W}(A)$ a je dán limitou $A = \lim_{i \to \infty} \mathcal{W}^i(B)$ pro libovolné $B \in H(\mathcal{M})$.

Důkaz. Viz. [2], str. 34 □

Definice 2.11. Buď $\mathcal{M}$ úplný metrický prostor, dále buď $D_i \subset \mathcal{M}$ pro $i = 1, ..., n$. Potom *segmentovaným systémem iterovaných funkcí (PIFS)* nazýváme množinu kontraktivních zobrazení $w_i : D_i \to \mathcal{M}$, pro $i = 1, ..., n$.

Poznámka 2.12. PIFS má (stejně jako v případě pro IFS) právě jeden pevný bod. Pokud budeme navíc uvažovat, že PIFS obsahuje i podmíněně kontraktivní, popřípadě neexpanzivní zobrazení, dá se ukázat, že i v tomto případě má toto PIFS právě jeden pevný bod. Viz. [3].

3. Algoritmus

Definice 3.1. Řekneme, že q je *podíl celočíselného dělení dvou celých čísel* $a \in \mathbb{N}_0, b \in \mathbb{N}$, pokud splňuje

$$a = b \cdot q + r\,, \qquad q \in \mathbb{N}_0, r \in \langle 0, |q|) \cap \mathbb{N}_0,$$

potom celočíselné dělení značíme $a \text{ div } b = q$ a zbytek po tomto celočíselném dělení značíme $a \bmod b = r$.

Tvrzení 3.2. *Mějme následující celočíselné dělení*

$$a \ div\ b, a \in \mathbb{N}_0, b \in \mathbb{N} - \{1\}.$$

Potom v $\mathbb{N}_0$ *s klasickou metrikou je toto celočíselné dělení neexpanzivní, a podmíněně kontraktivní.*

Definice 3.3. Definujme *zprávu* $\mathcal{M}$ jako konečnou posloupnost čísel z omezeného intervalu přirozených čísel. Délku zprávy (počet prvků této posloupnosti) označme $|\mathcal{M}|$. $N-tý$ prvek zprávy označme m_n.

Definice 3.4. Definujme *klíč* $\mathcal{K}$ jako konečnou posloupnost celočíselných dvojic $[\delta, \kappa]$, kde δ je z omezeného intervalu přirozených čísel a $\kappa \in \{2, 3, ..., 11\}$. Délku klíče (počet celočíselných dvojic této posloupnosti) označme $|\mathcal{K}|$. První prvek $n-té$ celočíselné dvojice klíče označme δ_n, druhý prvek této dvojice potom κ_n.

Algoritmus:
Mějme zprávu $\mathcal{A}$ délky $|\mathcal{A}|$ a klíč $\mathcal{K}$ délky $|\mathcal{K}|$. Dále mějme libovolnou zprávu $\mathcal{B}$ délky $|\mathcal{B}| = |\mathcal{A}|$.

1. Vytvořme novou (pomocnou) zprávu $\mathcal{E}$ délky $|\mathcal{E}| = |\mathcal{A}|$ podle následujícího předpisu

$$e_n = \left(\left(b_{((n+\delta_j-1) \bmod |\mathcal{E}|)+1}\right) \text{ div } \kappa_j\right) + a_n, \qquad j = ((n-1) \bmod |\mathcal{K}|) + 1. \quad (3.1)$$

2. Pokud

$$\sum_{i=1}^{|\mathcal{A}|} |b_i - e_i| = 0, \quad (3.2)$$

vrať $\mathcal{E}$ a ukonči výpočet. V opačném případě $\mathcal{B} = \mathcal{E}$ a jdi na bod #1.

Nyní jsme schopni rekonstruovat zprávu $\mathcal{A}$ podle následujícího předpisu

$$a_n = b_n - \left(\left(b_{((n+\delta_j-1) \bmod |\mathcal{E}|)+1}\right) \text{ div } \kappa_j\right), \qquad j = ((n-1) \bmod |\mathcal{K}|) + 1. \quad (3.3)$$

Poznámka 3.5. Z předpokladů plyne, že κ musí splňovat podmínku $\kappa \geq 2$. Avšak pokud vezmeme κ příliš velké, potom bude výsledek celočíselného dělení v algoritmu relativně malý. Tedy zašifrovaná zpráva by se příliš nelišila od zprávy původní. Proto v tomto algoritmu uvažujme $\kappa \in \{2, 3, ..., 11\}$.

Lemma 3.6. *Nechť $\mathcal{A}$ je zpráva, $\mathcal{K}$ je klíč a $\mathcal{E}$ je odpovídající zašifrovaná zpráva. Potom pro daný klíč $\mathcal{K}$ je zašifrovaná zpráva $\mathcal{E}$ jednoznačně určena, a je možno jí dosáhnout pouze z $\mathcal{A}$.*

Důkaz. Nechť $\mathcal{A}$, $\mathcal{B}$ jsou dvě navzájem odlišné zprávy stejné délky, a nechť $\mathcal{K}$ je libovolný klíč. Předpokládejme, že obě ze zpráv $\mathcal{B}$ a $\mathcal{A}$ mohou být zašifrovány stejným klíčem $\mathcal{K}$ do totožné zprávy $\mathcal{E}$. Potom pro každé n musí platit:

$$e_n = \left(\left(e_{((n+\delta_j-1) \bmod |\mathcal{E}|)+1}\right) \operatorname{div} \kappa_j\right) + a_n$$
$$e_n = \left(\left(e_{((n+\delta_j-1) \bmod |\mathcal{E}|)+1}\right) \operatorname{div} \kappa_j\right) + b_n,$$
$$j = ((n-1) \bmod |\mathcal{K}|) + 1.$$

Avšak z předpokladu $\mathcal{A} \neq \mathcal{B}$ plyne, že zde existuje alespoň jedno n takové, pro které platí $a_n \neq b_n$. Tudíž dostáváme spor. □

Příklad 3.7. Mějme zprávu $\mathcal{A} = \{50, 37, 85, 12, 69, 23, 52, 71, 49, 5\}$ délky $|\mathcal{A}| = 10$. Dále mějme klíč $\mathcal{K} = \{[35, 5], [9, 2], [73, 6]\}$ délky $|\mathcal{K}| = 3$ a zprávu $\mathcal{B} = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0\}$ délky $|\mathcal{B}| = 10$. Nyní vytvořme podle 3.1 novou zprávu $\mathcal{E}$ délky $|\mathcal{E}| = 10$.

$$\begin{aligned} e_1 &= ((b_6) \operatorname{div} 5) + a_1 = (0 \operatorname{div} 5) + 50 = 50 \\ e_2 &= ((b_1) \operatorname{div} 2) + a_2 = (0 \operatorname{div} 2) + 37 = 37 \\ &\vdots \\ e_{10} &= ((b_5) \operatorname{div} 5) + a_{10} = (0 \operatorname{div} 5) + 5 = 5 \end{aligned}$$

Dostali jsme tedy zprávu $\mathcal{E} = \{50, 37, 85, 12, 69, 23, 52, 71, 49, 5\}$. Položme $\mathcal{B} = \mathcal{E}$ a vypočítejme novou zprávu $\mathcal{E}$ následovně

$$\begin{aligned} e_1 &= ((b_6) \operatorname{div} 5) + a_1 = (23 \operatorname{div} 5) + 50 = 54 \\ e_2 &= ((b_1) \operatorname{div} 2) + a_2 = (50 \operatorname{div} 2) + 37 = 62 \\ e_3 &= ((b_6) \operatorname{div} 6) + a_3 = (23 \operatorname{div} 6) + 85 = 88 \\ &\vdots \\ e_{10} &= ((b_5) \operatorname{div} 5) + a_{10} = (69 \operatorname{div} 5) + 5 = 18 \end{aligned}$$

$$\begin{aligned} \mathcal{E} &= \{54, 62, 88, 21, 75, 31, 59, 97, 55, 18\} \\ &\vdots \\ \mathcal{E} &= \{56, 64, 90, 23, 79, 32, 64, 100, 59, 20\} \\ &\vdots \\ \mathcal{E} &= \{56, 65, 90, 23, 80, 32, 64, 103, 59, 20\} \\ &\vdots \\ \mathcal{E} &= \{56, 65, 90, 23, 80, 32, 65, 103, 59, 21\} \\ &\vdots \\ \mathcal{E} &= \{56, 65, 90, 23, 80, 32, 65, 103, 59, 21\} \end{aligned}$$

Po provedení šesté iterace si můžeme všimnout, že jsme dostali stejnou zprávu $\mathcal{E}$, jako v páté iteraci. Proto šifrování ukončíme. Nyní podle 3.3 vytvořme novou zprávu $\mathcal{A}'$

$$\begin{aligned} a'_1 &= b_1 - (b_6 \text{ div } 5) = 56 - (32 \text{ div } 5) = 56 - 6 = 50 \\ a'_2 &= b_2 - (b_1 \text{ div } 2) = 65 - (56 \text{ div } 2) = 65 - 28 = 37 \\ a'_3 &= b_3 - (b_6 \text{ div } 6) = 90 - (32 \text{ div } 6) = 90 - 5 = 85 \\ &\vdots \\ a'_{10} &= b_{10} - (b_5 \text{ div } 5) = 21 - (80 \text{ div } 5) = 21 - 16 = 5 \end{aligned}$$

Zpráva $\mathcal{A}' = \{50, 37, 85, 12, 69, 23, 52, 71, 49, 5\}$ je shodná s původní nezašifrovanou zprávou $\mathcal{A} = \{50, 37, 85, 12, 69, 23, 52, 71, 49, 5\}$.

V tomto příkladu si můžeme všimnout, že znaky zašifrované zprávy nabývají větších hodnot než znaky zprávy původní. V případě, že minimální hodnota v κ je rovna číslu 2, potom maximální hodnota výstupní abecedy může být až dvojnásobná oproti maximální hodnotě abecedy vstupní (to plyne z rovnice 3.1 pokud $((n + \delta_j - 1) \mod |\mathcal{E}|) + 1 = n$ a zároveň a_n je maximální hodnotou vstupní abecedy). To může působit problémy v případě, že jako zprávu k zašifrování bereme textovou zprávu převedenou do číselné reprezentace. Protože maximální hodnota výstupní abecedy může být až dvojnásobná, neměli bychom jak zpátky vyjádřit zašifrovanou zprávu v textové reprezentaci. Jedním způsobem je rozšířit původní abecedu o nové znaky (například podtržením původních znaků). Další možností je ponechat výstup v číselné reprezentaci (decimální, případně např. hexadecimální).

Poznámka 3.8. Pokud šifrujeme konstantní zprávu (nebo zprávu s opakující se sekvencí stejných znaků) krátkým klíčem ($|\mathcal{K}| \ll |\mathcal{M}|$), můžou se ve výsledné zašifrované zprávě vyskytovat opakující se sekvence stejných znaků (to plyne z rovnice 3.1 pokud je klíč krátký a a_n je konstantní). Pro zamezení tohoto jevu by se měl použít klíč alespoň stejné délky, jako je délka zprávy. To můžeme docílit například vhodným deterministickým algoritmem pro generování klíče potřebné délky z klíče originálního. Dalším přístupem může být definování klíče jako dvě různé konečné posloupnosti (Δ buď konečná posloupnost celých čísel z intervalu $\{2, 3, ..., 11\}$ a K buď konečná posloupnost z omezeného intervalu celých čísel), kde $|\Delta| = p$, $|K| = q$ a $p \neq q$. Označme $n - tý$ prvek K jako K_n a $n - tý$ prvek Δ jako Δ_n. Nyní je v rovnici 3.1 nutno pro e_n předefinovat $\kappa_j = K_{((n-1) \mod |K|)+1}$ a $\delta_j = \Delta_{((n-1) \mod |\Delta|)+1}$. Pak můžeme využít tento klíč pro zprávu délky nejvýše $p \times q$, aniž by se ve výsledné zašifrované zprávě vyskytovaly opakující se sekvence.

4. Závěr

Tento článek se zabýval popisem algoritmu FCBC šifry. Základní matematická teorie, popis algoritmu a jednoduchý příklad zde byly prezentovány. Samotná šifra FCBC je příbuzná k Vigenèrově šifře, ačkoli jsou jejich algoritmy odlišné. Vigenèrova šifra je symetrická substituční šifra, kde hodnota každého symbolu zprávy je posunuta o hodnotu znaku v klíči. V FCBC šifře se hodnoty znaků také posouvají, avšak o hodnotu znaku jiného (z téže zprávy), který je navíc modifikován

hodnotou znaku v klíči. Tudíž je výsledná zpráva v jistém smyslu soběpodobná. Nevýhodou této šifry může být větší maximální hodnota výstupní abecedy oproti abecedě vstupní. Navíc pro konstantní (nebo opakující se) zprávy je třeba klíč modifikovat (avšak stejný problém se vyskytuje i u Vigenèrovy šifry).

Softwarová implementace FCBC šifry, určena pouze pro ilustrativní účely, je ke stažení na adrese:

`http://fcbc.fme.vutbr.cz`

Síla FCBC šifry by měla být přibližně shodná se sílou Vigenèrovy šifry. Pro ověření tohoto tvrzení je třeba provést kryptoanalýzu FCBC. Touto tématikou by se autor rád zabýval ve svém dalším bádání.

Reference

[1] M. R. Barnsley: *Fractals Everywhere*, Academic Press, San Diego, 1988.
[2] Y. Fisher: *Fractal Image Compression: Theory and Application*, Springer Verlag, New York, 1995.
[3] S. K. Mitra, C. A. Murthy: *Mathematical framework to show the existence of attractor of partitioned iterative function systems*, Pattern Recognition **33** (2000), 859–869.

Tomáš Grísa, Ústav matematiky, Fakulta strojního inženýrství, Vysoké učení technické v Brně, Technická 2, 616 69 Brno, Česká republika,
e-mail: `y115559@stud.fme.vutbr.cz`

REPLIKÁTOROVÁ ROVNICE – PŘÍKLAD UŽITÍ MATEMATIKY V BIOLOGII

ZDENĚK POSPÍŠIL

Abstrakt. Přehledový článek uvádí matematický model vývoje společenstva populací zapsaný diferenciálními rovnicemi a model změny frekvence alel v jedné populaci zapsaný diferenční rovnicí. Tyto modely mají analogický tvar, což ukazuje, že z jistého hlediska jsou genetika a populační ekologie speciálními případy jedné obecnější teorie; konkrétně se jedná o evoluční syntézu (neodarwinismus). Tento výsledek ilustruje sílu matematiky také při popisu živé přírody.

Ve druhé části jsou uvedeny základní asymptotické vlastnosti sestavené rovnice.

1. Úvod

Matematika zasahuje do každého oboru lidské činnosti a biologie není v tomto směru výjimkou. Tato fráze je bezpochyby pravdivá, ovšem nic určitého neříká. Pokusíme se ji trochu zpřesnit. Užití matematiky může znamenat prostě nasazení výpočetní techniky pro uchování a analýzu experimentálních dat. Dlouhou tradici má také používání matematických modelů pro předpovídání vývoje nějakých procesů, a tím také pro jejich řízení; zakladatelskou prací v tomto směru je model šíření neštovic, který sestavil Daniel Bernoulli [2]. Takové použití matematiky však okamžitě vyvolává diskusi o adekvátnosti sestaveného modelu; to dokládá jednak bezprostřední reakce Jeana le Rond d'Alemberta na Bernoulliův článek [1], jednak diskuse pokračující dosud [5]. Novověká evropská přírodověda vidí v matematice „jazyk, kterým Bůh napsal přírodu" (Galileo Galilei, sr. diskusi možného významu této fráze v [18]). Jinak řečeno, matematika pomáhá porozumět přírodě, matematika je formou přírodních zákonů. Z tohoto pohledu je nejdokonalejší vědou fyzika, která své teorie formuluje rovnicemi; pojmy „matematická" a „teoretická" fyzika bývají zaměňovány, přestože jejich smysl se liší [17].

Na cestu přibližování se k ideálu fyziky se biologie vydala z Brna, kde Johann Gregor Mendel výsledky svých pokusů s křížením rostlin prováděných v klášterní zahradě zformuloval jako vzorečky o poměrech výskytu sledovaných znaků v následujících generacích [19]. Genetika po svém znovuobjevení na začátku dvacátého století cestu matematizace již neopustila. V této souvislosti je vhodné zmínit čistého matematika Godfreye Harolda Hardyho, který sice „neudělal nikdy nic užitečného" [13, str. 135 českého překladu], ale přesto se podle něho jmenuje základní

2010 *MSC.* Primární 92D15; Sekundární 92-01.

Klíčová slova. Replikátorová rovnice, matematický model, ekologie, populační genetika, evoluce.

Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

pravidlo genetiky (Hardyho-Weinbergův zákon, objevený dvakrát [12, 24] nezávisle na sobě). Další významnou postavou oboru je matematik a fyzik Ronald Aymler Fisher, který zkombinoval Mendelovy zákony s Darwinovou myšlenkou přirozeného výběru [9, 10]. Jeho teorii dále rozvíjeli biologové J. B. S. Haldane [11] a S. Wright [25]. Současný stav problematiky je zpracován např. v monografii [6].

Jinou matematicky propracovanou oblastí biologie je teorie přirozeného výběru na úrovni znaků, nikoliv genů. John Maynard Smith, původně letecký inženýr, použil spolu s Georgem Pricem terminologii teorie her k popisu konfliktů mezi zvířaty, darwinowskému „boji o život“ (struggle for existence), [22, 21]. Impuls Maynarda Smithe využili Peter Taylor a Leo Jonker a spojením některých myšlenek teorie her s diferenciálními rovnicemi vytvořili evoluční teorii her [23]. Jejich článek však zůstal téměř bez odezvy až do doby, kdy teorii použili Karl Sigmund a Peter Schuster [20] k upozornění na chybu ve slavné knize Richarda Dawkinse [4]. O evoluční teorii her se lze poučit v učebnici matematické biologie [3, kap. 4.] nebo podrobně v monografii [14].

Tento článek chce přiblížit zmíněné dva směry teoretické biologie – genetiku a evoluci. V první části následující sekce je sestavena rovnice popisující změny frekvence alel (variant genů) v populaci, v její druhé části rovnice popisující vývoj struktury nějakého společenstva biologických (pseudo)druhů. Základním výsledkem je to, že oba procesy popisuje stejná rovnice, kterou nazýváme replikátorová (replicator equation). To ukazuje, že se v jistém smyslu jedná o proces jediný (podnětnou diskusi o „fundamentální jednotě přírody“ vyjádřené tím, že různé děje jsou popsány stejnými rovnicemi, uvádí R. P. Feynman [8, díl 3, kap. 12.7]) a je hlubším teoretickým zdůvodněním neodarwinismu – syntézy evoluce a genetiky.

Ve třetí sekci jsou studovány některé základní asymptotické vlastnosti řešení replikátorové rovnice. Zejména je charakterizováno asymptoticky stabilní stacionární řešení, tzv. evolučně stabilní stav, které může být vhodnou formalizací pojmu „zamrzlá evoluce“, o níž mluví provokativně psaná kniha [7]. Základní Věta 3.4 o řešení rovnice (2.14) byla původně dokázána v článku [15] pro speciálnější rovnici (3.1).

Používaná symbolika je standardní. Méně obvyklý je snad jen Hadamardův součin vektorů; pro vektory $\boldsymbol{x} = (x_1, x_2, \dots, x_k), \boldsymbol{y} = (y_1, y_2, \dots, y_k)$ je definován vztahem $\boldsymbol{x} \circ \boldsymbol{y} = (x_1 y_1, x_2 y_2, \dots, x_k y_k)$. Potřebné poznatky o obyčejných diferenciálních rovnicích lze najít např. v učebnici [16].

2. Konstrukce replikátorové rovnice

2.1. Genetika populací

Populační genetika popisuje přenos genů (nosičů dědičnosti) mezi generacemi nějakých organismů. V tomto článku se budeme věnovat jednomu velice zjednodušenému případu zmíněného přenosu. Představme si, že dospělí jedinci nějakého druhu produkují pohlavní buňky, *gamety.* Spojením dvou gamet vznikne nový jedinec, *zygota.* Ten může dospět do plodného věku, produkovat gamety a celý cyklus se bude opakovat. Budeme předpokládat, že čas, který uplyne od stadia gamety přes zygotu a plodného jedince do produkce dalších gamet, je jednotkový,

tj. pokud gamety první, *parentální,* generace jsou vytvořeny v čase t, pak gamety následující, *první filiální,* generace jsou vytvořeny v čase $t+1$.

Geny si budeme představovat mendelovským způsobem, podrobnosti na molekulární úrovni nejsou pro potřeby tohoto článku relevantní. Jeden gen je představován nějakým místem na chromozomu, *lokusem.* Zygoty a dospělí jedinci mají chromozomy v párech, jsou *diploidní.* To znamená, že na jednom lokusu jsou dvě varianty genetického materiálu, *alely,* které nemusí být shodné. Tato neuspořádaná dvojice alel určuje *genotyp* jedince. Naproti tomu gamety mají každý chromozom jen jednou, jsou *haploidní,* a tedy nesou jen jednu alelu.

Budeme uvažovat chromozom, který není pohlavní, a na něm jeden lokus se dvěma možnými alelami, které označíme A, a. Jako první předpoklad přijmeme, že do gamet přechází alely náhodně. To znamená, že gameta vyprodukovaná jedincem s genotypem Aa s pravděpodobností $\frac{1}{2}$ obsahuje alelu A a se stejnou pravděpodobností $\frac{1}{2}$ obsahuje alelu a. Gameta vyprodukovaná jedincem genotypu AA jistě, tj. s pravděpodobností 1, obsahuje alelu A, gameta vyprodukovaná jedincem genotypu aa jistě obsahuje alelu a.

Označme $x(t)$ podíl gamet, které nesou alelu A, mezi všemi gametami v čase t. Hodnotu $x(t)$ můžeme také považovat za klasickou pravděpodobnost, že náhodně vybraná gameta nese alelu A. V důsledku toho je podíl gamet, které mezi všemi gametami v čase t nesou alelu a, roven $1-x(t)$. Budeme také předpokládat, že tyto podíly jsou stejné i u samotných samičích gamet, nebo samotných samčích gamet. Jinak řečeno, subpopulace samic a samců jsou z hlediska uvažovaného lokusu geneticky identické.

Budeme dále předpokládat, že v populaci dochází k náhodnému křížení, že populace je *panmiktická.* To znamená, že gamety se při vytváření zygot spojují náhodně a nezávisle na alelách, které nesou; libovolná samičí gameta se může spojit s libovolnou samčí gametou. (Tento proces je asi lepší si představovat jako pylová zrna nesená náhodně vanoucími větry na blizny náhodně rozmístěné v krajině, ne jako spermie a vajíčka spojující se v těle samice.) Označme Z_{AA}, Z_{Aa} a Z_{aa} počet zygot genotypu AA, Aa a aa. Položme $Z = Z_{AA} + Z_{Aa} + Z_{aa}$. Dále označme

$$p_{AA} = \frac{Z_{AA}}{Z}, \quad p_{Aa} = \frac{Z_{Aa}}{Z}, \quad p_{aa} = \frac{Z_{aa}}{Z} \tag{2.1}$$

podíl zygot příslušného genotypu mezi všemi zygotami, který můžeme považovat za klasickou pravděpodobnost, že náhodně vybraná zygota je daného genotypu. Zygota genotypu AA je realizací nezávislých náhodných jevů, že gameta od jednoho rodiče nese alelu A, pravděpodobnost tohoto jevu je rovna $x(t)$, a že gameta od druhého rodiče nese také alelu A, což je jev se stejnou pravděpodobností $x(t)$. Pravděpodobnost vzniku zygoty genotypu AA je tedy rovna $p_{AA} = x(t)x(t) = x(t)^2$. Analogickou úvahou zjistíme, že $p_{aa} = \big(1-x(t)\big)^2$. Poněvadž každá zygota je právě jednoho z možných genotypů, platí $p_{Aa} = 1-(p_{AA}+p_{aa})$. Dostáváme tak vztahy

$$p_{AA} = \frac{Z_{AA}}{Z} = x(t)^2, \quad p_{Aa} = \frac{Z_{Aa}}{Z} = 2x(t)\big(1-x(t)\big), \quad p_{aa} = \frac{Z_{aa}}{Z} = \big(1-x(t)\big)^2. \tag{2.2}$$

Budeme předpokládat, že zygoty různých genotypů mohou mít různou pravděpodobnost, že se dožijí plodného věku, a že plodní jedinci různých genotypů

mohou mít různou plodnost, tj. vyprodukují různý počet gamet. Jinak řečeno, výběr (ať už přirozený nebo umělý) působí na úrovni genotypu. Označme S_{AA} podíl zygot mezi všemi zygotami genotypu AA, které dosáhnou plodné fáze, tj. pravděpodobnost, že se zygoty genotypu AA dožijí plodnosti. Dále označme m_{AA} počet gamet, které vyprodukuje dospělý jedinec genotypu AA. V analogickém významu budeme používat označení S_{Aa}, S_{aa}, m_{Aa} a m_{aa}. (Poznámka pro ty, kteří si pamatují ze střední školy o genetice něco více: Genotyp „není vidět“, proto selekce na úrovni genotypu nemůže probíhat. Ovšem fenotyp je genotypem jednoznačně určen. Pokud by alela A byla dominantní a alela a recesivní, platilo by $S_{AA} = S_{Aa}$, $m_{AA} = m_{Aa}$.)

Označme dále N_{AA}, N_{Aa}, N_{aa} počet plodných jedinců příslušných genotypů, a M_{AA}, M_{Aa} a M_{aa} celkový počet gamet, které vyprodukují všichni jedinci příslušného genotypu. Pak s využitím vztahů (2.1) a (2.2) dostaneme

$$M_{AA} = m_{AA}N_{AA} = m_{AA}S_{AA}Z_{AA} = m_{AA}S_{AA}Zx(t)^2 \tag{2.3}$$

a podobně

$$M_{Aa} = 2m_{Aa}S_{Aa}Zx(t)\big(1 - x(t)\big), \quad M_{aa} = m_{aa}S_{aa}Z\big(1 - x(t)\big)^2. \tag{2.4}$$

Všechny zygoty genotypu AA, kterých bylo Z_{AA}, můžeme považovat za „prvotní producenty“ celkem $M_{AA} = m_{AA}S_{AA}Z_{AA}$ gamet. To znamená, že jednu zygotu genotypu AA můžeme považovat za původce

$$\frac{M_{AA}}{Z_{AA}} = m_{AA}S_{AA}$$

gamet. Tuto veličinu lze považovat za *reprodukční zdatnost* genotypu AA. Označme ji w_{AA}. Stejnou úvahu můžeme provést a analogické označení zavést pro ostatní genotypy. Pak přepíšeme vztahy (2.3), (2.4) ve stručnějším tvaru

$$M_{AA} = w_{AA}Zx(t)^2, \quad M_{Aa} = 2w_{Aa}Zx(t)\big(1 - x(t)\big), \quad M_{aa} = w_{aa}Z\big(1 - x(t)\big)^2. \tag{2.5}$$

Všechny gamety vyprodukované jedinci genotypu AA nesou alelu A a polovina gamet vyprodukovaných jedinci genotypu Aa nese také alelu A. Celkový počet K_A gamet s alelou A v čase $t + 1$ tedy je $K_A = M_{AA} + \frac{1}{2}M_{Aa}$. Podobně celkový počet K_a gamet s alelou a je $K_a = M_{aa} + \frac{1}{2}M_{Aa}$. Pro podíl gamet nesoucích alelu A v čase $t + 1$ nyní s využitím vztahů (2.5) dostaneme vyjádření

$$x(t+1) = \frac{K_A}{K_A + K_a} = \frac{w_{AA}x(t)^2 + w_{Aa}x(t)\big(1 - x(t)\big)}{w_{AA}x(t)^2 + 2w_{Aa}x(t)\big(1 - x(t)\big) + w_{aa}\big(1 - x(t)\big)^2},$$

z něhož bezprostředně plyne *Fischerova-Haldaneova-Wrightova rovnice populační genetiky*

$$x(t+1) = \frac{w_{AA}x(t) + w_{Aa}\big(1 - x(t)\big)}{w_{AA}x(t)^2 + 2w_{Aa}x(t)\big(1 - x(t)\big) + w_{aa}\big(1 - x(t)\big)^2}x(t). \tag{2.6}$$

Veličiny w_{AA}, w_{Aa} a w_{aa} vyjadřují reprodukční zdatnosti jednotlivých genotypů. Nyní můžeme uvažovat náhodnou veličinu „reprodukční zdatnost populace“

a vypočítat její střední hodnotu $\bar{w}$ pomocí pravděpodobnostní funkce genotypů (2.2):

$$\begin{aligned}\bar{w} &= w_{AA}p_{AA} + w_{Aa}p_{Aa} + w_{aa}p_{aa} = \\ &= w_{AA}x(t)^2 + 2w_{Aa}x(t)\big(1-x(t)\big) + w_{aa}\big(1-x(t)\big)^2 = \\ &= \Big(w_{AA}x(t) + w_{Aa}\big(1-x(t)\big)\Big)x(t) + \Big(w_{Aa}x(t) + w_{aa}\big(1-x(t)\big)\Big)\big(1-x(t)\big).\end{aligned} \tag{2.7}$$

Při označení

$$w_A = w_{AA}x(t) + w_{Aa}\big(1-x(t)\big), \quad w_a = w_{Aa}x(t) + w_{aa}\big(1-x(t)\big) \tag{2.8}$$

můžeme psát

$$\bar{w} = w_A x(t) + w_a\big(1-x(t)\big). \tag{2.9}$$

Střední hodnotu reprodukční zdatnosti populace tedy vyjadřuje výraz, který je formálně roven střední hodnotě jisté náhodné veličiny na dvouprvkovém pravděpodobnostním prostoru. Tato náhodná veličina má hodnotu w_A na jevu, jehož pravděpodobnost je rovna $x(t)$, tedy na jevu, že náhodně vybraná alela je A. Analogickou úvahu provedeme pro w_a. Hodnoty w_A, w_a tedy můžeme interpretovat jako reprodukční zdatnosti alely A a alely a.

Základní rovnici (2.6) můžeme s označením (2.7), (2.8) a (2.9) přepsat ve tvaru

$$x(t+1) = \frac{w_A}{\bar{w}}.$$

Podíl

$$f_A = \frac{w_A}{\bar{w}}$$

vyjadřuje relativní reprodukční zdatnost alely A vzhledem k reprodukční zdatnosti populace. Hodnota f_A samozřejmě závisí na relativní frekvenci x alely A, $f_A = f_A(x)$. Fischerovu-Haldaneovu-Wrightovu rovnici (2.6) nyní můžeme vyjádřit ve tvaru rekurentní formule

$$x(t+1) = f_A\big(x(t)\big)x(t)$$

nebo autonomní diferenční rovnice

$$\Delta x = x\big(f_A(x) - 1\big). \tag{2.10}$$

Tuto rovnici můžeme přečíst: „pokud je hodnota relativní reprodukční zdatnosti alely A větší než 1, pak relativní zastoupení alely A v populaci roste.“

2.2. Vývoj společenstva populací

Nejprve budeme modelovat vývoj velikosti jedné populace. Budeme ji považovat za homogenní, tj. takovou, že všichni jedinci tvořící populaci jsou stejní. Označme $n = n(t)$ velikost populace v čase t. Tuto velikost můžeme vyjádřit např. počtem jedinců, počtem jedinců vztažených na jednotku plochy (populační hustotou), celkovou biomasou populace a podobně. Budeme předpokládat, že populace je velká, takže funkci n budeme považovat za diferencovatelnou. Velikost populace $n(t+h)$ po krátkém časovém úseku délky h bude zřejmě rovna její velikosti na počátku uvažovaného úseku zvětšené o množství nově narozených jedinců a zmenšené o množství jedinců uhynulých. Abychom tuto myšlenku vyjádřili přesněji,

označíme symbolem d poměr uhynulých jedinců ku všem za jednotku času, tzv. *úmrtnost (death rate)* populace, a b množství novorozenců, které „vyprodukuje“ jedinec za jednotku času, tzv. *porodnost (birth rate)*. Při tomto označení máme

$$n(t+h) = n(t) + bn(t)h - dn(t)h.$$

Pro zjednodušení zápisu položíme $r = b-d$, rovnost obvyklým způsobem upravíme a provedeme limitní přechod $h \to 0$. Dostaneme obyčejnou diferenciální rovnici

$$n' = rn.$$

Tato rovnice je lineární homogenní pouze v případě, že *růstový koeficient (growth rate)* r nezávisí na velikosti populace n. Tak tomu však ve většině případů není – v malé populaci může být obtížné najít partnera k páření (je malá porodnost), velká populace spotřebuje zdroje prostředí (je velká úmrtnost) a podobně. Je tedy $r = \varphi(n)$ a rovnice růstu populace nabývá tvar

$$n' = n\varphi(n). \tag{2.11}$$

Uvažme nyní společenstvo tvořené k populacemi. Velikosti jednotlivých populací označíme $n_i = n_i(t)$ a jejich růstové koeficienty φ_i, $i = 1, 2, \ldots, k$. Jednotlivé populace se budou vyvíjet podle rovnic analogických (2.11), ovšem velikosti růstových koeficientů mohou záviset na velikostech jednotlivých populací – populace si mohou konkurovat ve využívání zdrojů prostředí, jedna může být potravou jiné a podobně. Dostáváme tedy model společenstva ve tvaru autonomního systému k obyčejných diferenciálních rovnic

$$n_i' = n_i\varphi_i(n_1, n_2, \ldots, n_k), \quad i = 1, 2, \ldots, k,$$

nebo v zápisu využívajícím označení $\boldsymbol{n} = (n_1, n_2, \ldots, n_k)$

$$n_i' = n_i\varphi_i(\boldsymbol{n}), \quad i = 1, 2, \ldots, k, \tag{2.12}$$

případně jako jednu vektorovou rovnici

$$\boldsymbol{n}' = \boldsymbol{n} \circ \boldsymbol{\varphi}(\boldsymbol{n}),$$

kde $\circ$ označuje Hadamardův součin vektorů (součin „po složkách“).

Označme nyní

$$N = N(t) = \sum_{i=1}^{k} n_i(t), \quad x_i = x_i(t) = \frac{n_i(t)}{N(t)},\ i = 1, 2, \ldots, k$$

celkovou velikost společenstva a relativní zastoupení jednotlivých populací ve společenstvu. Pak vzhledem k (2.12) platí

$$x_i' = \frac{n_i'N - n_iN'}{N^2} = \frac{n_i'}{N} - \frac{n_i}{N}\sum_{j=1}^{k}\frac{n_j'}{N} = \frac{n_i}{N}\varphi_i(\boldsymbol{n}) - \frac{n_i}{N}\sum_{j=1}^{k}\frac{n_j}{N}\varphi_j(\boldsymbol{n}) =$$

$$= \frac{n_i}{N}\left(\varphi_i(\boldsymbol{n}) - \sum_{j=1}^{k}\frac{n_j}{N}\varphi_j(\boldsymbol{n})\right).$$

Při označení

$$f_i(\boldsymbol{x}) = \varphi_i(N\boldsymbol{x}) = \varphi_i(\boldsymbol{n})$$

dostaneme systém obyčejných diferenciálních rovnic

$$x_i' = x_i \left(f_i(\boldsymbol{x}) - \sum_{j=1}^{k} x_j f_j(\boldsymbol{x}) \right), \quad i = 1, 2, \ldots, k. \tag{2.13}$$

Poněvadž $\sum_{j=1}^{k} x_j = 1$, lze druhý výraz v závorce považovat za vážený průměr jednotlivých f_j, $j = 1, 2, \ldots, k$. Proto ho označíme $\bar{f}$ a systém (2.13) přepíšeme ve tvaru

$$x_i' = x_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right), \quad i = 1, 2, \ldots, k. \tag{2.14}$$

Tuto rovnici můžeme přečíst: „pokud je hodnota veličiny f pro i-tou populaci větší, než průměrná, pak relativní zastoupení této populace ve společenstvu roste.“ Tato interpretace ukazuje, že veličina f_i vyjadřuje *darwinovskou zdatnost (fitness)* i-té populace. Nějaká přirozená jednotka darwinovské zdatnosti neexistuje. Můžeme ji tedy zvolit tak, aby průměrná zdatnost $\bar{f}$ byla jednotková; toto tvrzení budeme precizovat v následující sekci, Lemma 3.2. Diferenciální rovnice (2.14) je pak bezprostřední spojitou analogií diferenční rovnice (2.10).

Rovnice (2.10) modeluje vývoj relativního zastoupení alely v populaci; alela je jakási entita, která se stále znovu vytváří jako kopie entity rodičovské, replikuje se v čase. Rovnice (2.14), přesněji systém obyčejných diferenciálních rovnic, modeluje vývoj relativního zastoupení jedinců jednoho druhu ve společenstvu; ti také stále znovu vznikají jako kopie svých rodičů, replikují se v čase. Z těchto důvodů se (2.14) nazývá *replikátorová rovnice.* Systém rovnic (2.14) můžeme totiž zapsat jako jednu vektorovou rovnici

$$\boldsymbol{x}' = \boldsymbol{x} \circ \boldsymbol{f}(\boldsymbol{x}) - (\boldsymbol{x} \cdot \boldsymbol{f}(\boldsymbol{x}))\, \boldsymbol{x},$$

kde $\boldsymbol{f} = (f_1, f_2, \ldots, f_k)$, symboly $\circ$ a $\cdot$ označují Hadamardův a skalární součin vektorů.

3. Vlastnosti replikátorové rovnice

Budeme předpokládat, že všechny funkce f_i, $i = 1, 2, \ldots, k$, jsou spojitě diferencovatelné. Pak mají pravé strany systému (2.14) ohraničené parciální derivace podle všech proměnných, což znamená, že jsou Lipschitzovské v proměnné $\boldsymbol{x}$ a počáteční úloha pro systém (2.14) je jednoznačně řešitelná na intervalu $[0, \infty)$.

Věta 3.1. *Nechť* $\boldsymbol{x} = \boldsymbol{x}(t) = \big(x_1(t), x_2(t), \ldots, x_k(t)\big)$ *je řešení systému* (2.14). *Pak platí*

(i) *Je-li* $\sum_{i=1}^{k} x_i(0) = 1$, *pak je* $\sum_{i=1}^{k} x_i(t) = 1$ *pro každé* $t \geq 0$.

(ii) *Je-li* $x_i(0) = 0$, *resp.* $x_i(0) > 0$, *pro nějaké* $i \in \{1, 2, \ldots, k\}$, *pak je* $x_i(t) = 0$, *resp.* $x_i(t) > 0$, *pro všechna* $t \geq 0$.

Důkaz: Položme $y(t) = \sum_{i=1}^{k} x_i(t) - 1$. Pak

$$\frac{\mathrm{d}}{\mathrm{d}t} y = \sum_{i=1}^{k} x_i' = \sum_{i=1}^{k} x_i \big(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \big) = \bar{f}(\boldsymbol{x}) - \left(\sum_{i=1}^{k} x_i \right) \bar{f}(\boldsymbol{x}) = 0.$$

To znamená, že $y(t) \equiv const$, podrobněji

$$y(t) \equiv y(0) = \sum_{i=1}^{k} x_i(0) - 1$$

a odtud plyne tvrzení (i).

Systém (2.14) má podle předpokladu jediné řešení, jeho i-tá složka splňuje rovnost

$$x_i(t) = x_i(0) \exp \int_0^t x_i(s)\Big(f_i\big(\boldsymbol{x}(s)\big) - \bar{f}\big(\boldsymbol{x}(s)\big)\Big)\mathrm{d}s,$$

ze které plynou obě tvrzení (ii). □

Větu můžeme interpretovat tak, že složky řešení replikátorové rovnice vyjadřují časově závislé relativní zastoupení jednotlivých komponent modelovaného biologického systému – alel v genotypu nebo druhů ve společenstvu. Komponenta, která není v systému od začátku, se v něm v průběhu času neobjeví; v systému nic nového nevzniká. Replikátorová rovnice tedy popisuje deterministickou část evoluce, tj. selekci, nikoliv složku stochastickou, tj. mutace. Tento závěr není nějak překvapivý, replikátorová rovnice byla takto sestavena. Druhá z vlastností (ii) říká, že komponenta systému, který v něm je na začátku, v konečném čase nezmizí. Nevylučuje však možnost

$$\lim_{t\to\infty} x_i(t) = 0$$

pro složku takovou, že $x_i(0) > 0$. Komponenta systému může vymizet (vyhynout) v „dostatečně dlouhém“ čase.

Připomeňme, že množina

$$S_k = \left\{ \boldsymbol{x} = (x_1, x_2, \ldots, x_k) \in \mathbb{R}^k : \sum_{i=1}^{k} x_i = 1, (\forall i) x_i \geq 0 \right\}$$

se nazývá *k-rozměrný simplex.* Větu 3.1 nyní můžeme přeformulovat: simplex S_k, jeho vnitřek S_k° a hranice ∂S_k jsou pozitivně invariantní množiny systému (2.14).

Lemma 3.2. *Buď $\Psi : S_k \to \mathbb{R}$ spojitě diferencovatelná funkce. Položme*

$$g_i = g_i(\boldsymbol{x}) = f_i(\boldsymbol{x}) + \Psi(\boldsymbol{x})$$

pro $i \in \{1, 2, \ldots, k\}$. Funkce $\boldsymbol{x} = \boldsymbol{x}(t)$ je řešením rovnice (2.14) *právě tehdy, když je řešením rovnice*

$$x_i' = x_i \left(g_i(\boldsymbol{x}) - \sum_{j=1}^{k} x_j g_j(\boldsymbol{x}) \right).$$

Důkaz: Tvrzení plyne z rovnosti

$$x_i' = x_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x})\right) = x_i \left(f_i(\boldsymbol{x}) - \sum_{j=1}^{k} x_j f_j(\boldsymbol{x}) \right) =$$

$$= x_i \left(g_i(\boldsymbol{x}) - \Psi(\boldsymbol{x}) - \sum_{j=1}^{k} x_j \big(g_j(\boldsymbol{x}) - \Psi(\boldsymbol{x})\big) \right) = x_i \left(g_i(\boldsymbol{x}) - \sum_{j=1}^{k} x_j g_j(\boldsymbol{x}) \right).$$

□

Volbou $\Psi(\boldsymbol{x}) = 1 - \bar{f}(\boldsymbol{x})$ dostaneme

$$\sum_{j=1}^{k} x_j g_j(\boldsymbol{x}) = \sum_{j=1}^{k} x_j \left(f_j(\boldsymbol{x}) + 1 - \bar{f}(\boldsymbol{x})\right) = \bar{f}(\boldsymbol{x}) + 1 - \bar{f}(\boldsymbol{x}) = 1.$$

Průměrnou darwinovskou zdatnost společenstva tedy můžeme bez újmy na obecnosti pokládat za jednotkovou.

Lemma 3.3 (Speciální tvar Jensenovy nerovnosti)**.** *Buď φ diferencovatelná ryze konvexní funkce definovaná na intervalu $I \subseteq \mathbb{R}$, tj. pro všechna $\xi_1, \xi_2 \in I$ platí*

$$\varphi(\xi_1) \geq \varphi(\xi_2) + \varphi'(\xi_2)(\xi_2 - \xi_1)$$

a nerovnost je ostrá, kdykoliv $\xi_1 \neq \xi_2$. Pak pro libovolná $\xi_1, \xi_2, \ldots, \xi_k \in I$ a každý vektor $\boldsymbol{p} = (p_1, p_2, \ldots, p_k) \in S_k^\circ$ platí

$$\varphi\left(\sum_{i=1}^{k} p_i \xi_i\right) \leq \sum_{i=1}^{k} p_i \varphi(\xi_i),$$

přičemž rovnost nastane právě tehdy, když $\xi_1 = \xi_2 = \cdots = \xi_k$.

Důkaz: Z konvexnosti funkce φ plyne, že pro libovolný index $j \in \{1, 2, \ldots, k\}$ platí

$$\varphi(\xi_j) \geq \varphi\left(\sum_{i=1}^{k} p_i \xi_i\right) + \varphi'\left(\sum_{i=1}^{k} p_i \xi_i\right)\left(\xi_j - \sum_{i=1}^{k} p_i \xi_i\right).$$

Každou takovou nerovnost vynásobíme p_j a sečteme je přes všechny indexy j. Dostaneme

$$\begin{aligned}
&\sum_{j=1}^{k} p_j \varphi(\xi_j) \geq \\
&\geq \sum_{j=1}^{k} p_j \varphi\left(\sum_{i=1}^{k} p_i \xi_i\right) + \sum_{j=1}^{k} p_j \left[\xi_j \varphi'\left(\sum_{i=1}^{k} p_i \xi_i\right) - \varphi'\left(\sum_{i=1}^{k} p_i \xi_i\right)\left(\sum_{i=1}^{k} p_i \xi_i\right)\right] = \\
&= \varphi\left(\sum_{i=1}^{k} p_i \xi_i\right) + \varphi'\left(\sum_{i=1}^{k} p_i \xi_i\right) \sum_{j=1}^{k} p_j \xi_j - \sum_{j=1}^{k} p_j \varphi'\left(\sum_{i=1}^{k} p_i \xi_i\right) \sum_{i=1}^{k} p_i \xi_i = \\
&= \varphi\left(\sum_{i=1}^{k} p_i \xi_i\right).
\end{aligned}$$

Pokud ne všechna ξ_j jsou stejná, pak pro j takové, že $\xi_j = \max\{\xi_1, \xi_2, \ldots, \xi_k\}$, platí $\xi_j > \sum_{i=1}^{k} p_i \xi_i$, neboť na pravé straně je vážený průměr čísel $\xi_1, \xi_2, \ldots, \xi_k$. V takovém případě je nerovnost ostrá. □

Věta 3.4. *Nechť pro $\hat{\boldsymbol{x}} = (\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k) \in S_k$ existuje okolí $U \subseteq S_k$ takové, že pro všechna $\boldsymbol{x} \in U \setminus \{\hat{\boldsymbol{x}}\}$ platí*

$$\sum_{i=1}^{k} \hat{x}_i f_i(\boldsymbol{x}) > \bar{f}(\boldsymbol{x}).$$

Pak $\hat{\boldsymbol{x}}$ je asymptoticky stabilní stacionární bod systému (2.14).

Důkaz: Pro $\boldsymbol{y} \in S_k$ označme $\operatorname{supp} \boldsymbol{y} = \{i \in \{1, 2, \dots, k\} : y_i > 0\}$ a pro vektor $\boldsymbol{x} = (x_1, x_2, \dots, x_k)$ položme

$$V(\boldsymbol{x}) = \prod_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i^{\hat{x}_i} - \prod_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} x_i^{\hat{x}_i}.$$

Ukážeme, že V je ljapunovskou funkcí systému (2.14) v bodě $\hat{\boldsymbol{x}}$.

Podle Lemmatu 3.3 pro bod $\boldsymbol{x}$ z okolí bodu $\hat{\boldsymbol{x}}$ takový, že $\operatorname{supp} \boldsymbol{x} = \operatorname{supp} \hat{\boldsymbol{x}}$ platí

$$\sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \ln \frac{\hat{x}_i}{x_i} = \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \left(- \ln \frac{x_i}{\hat{x}_i} \right) \geq - \ln \left(\sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \frac{x_i}{\hat{x}_i} \right) = - \ln 1 = 0.$$

To znamená, že

$$\sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \ln \hat{x}_i \geq \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \ln x_i, \quad \text{tj.} \quad \ln \left(\prod_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i^{\hat{x}_i} \right) \geq \ln \left(\prod_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} x_i^{\hat{x}_i} \right).$$

Odlogaritmováním poslední nerovnosti dostaneme $V(\boldsymbol{x}) \geq 0$. Rovnost přitom nastane právě tehdy, když

$$\frac{x_i}{\hat{x}_i} = \frac{x_j}{\hat{x}_j} \quad \text{pro všechny indexy } i, j \in \operatorname{supp} \hat{\boldsymbol{x}},$$

tj. právě tehdy, když $\boldsymbol{x} = c\hat{\boldsymbol{x}}$ pro nějakou konstantu c. Poněvadž $\boldsymbol{x}, \hat{\boldsymbol{x}} \in S_k$, musí být $c = 1$. Rovnost $V(\boldsymbol{x}) = 0$ tedy nastává právě tehdy, když $\boldsymbol{x} = \hat{\boldsymbol{x}}$.

Označme nyní

$$P(\boldsymbol{x}) = \prod_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} x_i^{\hat{x}_i}.$$

Pak je $P(\boldsymbol{x}) > 0$ a $V(\boldsymbol{x}) = P(\hat{\boldsymbol{x}}) - P(\boldsymbol{x})$. Odtud plyne, že pro derivaci funkce V vzhledem k rovnici (2.14) platí $\dot{V}(\boldsymbol{x}) = -\dot{P}(\boldsymbol{x})$. Dále podle předpokladu věty platí

$$\begin{aligned} \frac{\dot{P}(\boldsymbol{x})}{P(\boldsymbol{x})} = \frac{\mathrm{d}}{\mathrm{d}t} \ln P(\boldsymbol{x}) = \frac{\mathrm{d}}{\mathrm{d}t} \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \ln x_i = \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \frac{x_i'}{x_i} = \\ = \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i \left(f_i(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right) = \sum_{i \in \operatorname{supp} \hat{\boldsymbol{x}}} \hat{x}_i f_i(\boldsymbol{x}) - \hat{f}(\boldsymbol{x}) > 0. \end{aligned}$$

Odtud plyne $\dot{V}(\boldsymbol{x}) < 0$. □

Vektor $\hat{\boldsymbol{x}}$ splňující předpoklady Věty 3.4 vyjadřuje takovou strukturu modelovaného biologického systému, při níž se relativní zastoupení jednotlivých komponent nemění. Pokud je struktura systému „blízko“ struktuře $\hat{\boldsymbol{x}}$, tj. struktura se „příliš neliší“ od $\hat{\boldsymbol{x}}$, pak se systém do stavu $\hat{\boldsymbol{x}}$ vyvine a zůstane v něm. To nás opravňuje nazvat $\hat{\boldsymbol{x}}$ splňující předpoklady věty *evolučně stabilním stavem vzhledem ke zdatnostem* $f_1, f_2, \dots, f_k$.

Uvažujme nyní jednodušší, ale důležitý případ, kdy zdatnosti $f_1, f_2, \dots, f_k$ jsou lineárními funkcemi vektoru $\boldsymbol{x}$, tj. kdy existují konstanty a_{ij}, $i, j = 1, 2, \dots, k$

takové, že

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{k} a_{ij} x_j, \quad i = 1, 2, \ldots, k.$$

Nechť A je matice s prvky a_{ij}, $\mathsf{A} = \left(a_{ij}\right)_{i,j=1}^{k}$, vektory budeme považovat za sloupcové. Pak $f_i(\boldsymbol{x}) = (\mathsf{A}\boldsymbol{x})_i$ (i-tá složka vektoru $\mathsf{A}\boldsymbol{x}$) a

$$\bar{f}(\boldsymbol{x}) = \sum_{i=1}^{k} \sum_{j=1}^{k} x_i a_{ij} x_j = \boldsymbol{x}^{\mathsf{T}} \mathsf{A} \boldsymbol{x}.$$

Replikátorovou rovnici (2.14) můžeme v takovém případě zapsat ve tvaru

$$x_i' = x_i \left((\mathsf{A}\boldsymbol{x})_i - \boldsymbol{x}^{\mathsf{T}} \mathsf{A} \boldsymbol{x} \right), \quad i = 1, 2, \ldots, k, \tag{3.1}$$

nebo jako jednu vektorovou rovnici $\boldsymbol{x}' = \boldsymbol{x} \circ \mathsf{A}\boldsymbol{x} - \left(\boldsymbol{x}^{\mathsf{T}} \mathsf{A} \boldsymbol{x}\right) \boldsymbol{x}$. V této podobě – s lineárními zdatnostmi – byla replikátorová rovnice zavedena v článcích [23, 20].

Matice A reprezentuje nějakou maticovou hru. Jednotlivé druhy tvořící společenstvo (nebo jednotlivé alely na vybraném lokusu) lze v tomto pojetí interpretovat jako ryzí strategie, které používají hráči v „boji o život“. Zdatnosti $f_i(\boldsymbol{x}) = \sum_{i=1}^{k} a_{ij} x_j$ pak představují střední výhru i-tého druhu při konfliktech s ostatními druhy, se kterými se náhodně potkává; pravděpodobnost, že se setká s j-tým druhem, je totiž rovna relativnímu výskytu j-tého druhu ve společenstvu. Tato úvaha umožňuje (aspoň v prvním přiblížení) stanovit zdatnosti jednotlivých druhů.

Reference

[1] J. d'Alembert: *Onzième mémoire, Sur l'application du calcul des probabilités à l'inoculation de la petite vérole*, Opuscules mathématiques **2** (1761), 26–95.

[2] D. Bernoulli: *Essai d'un nouvelle analyse de la mortalité causée par la petite vérole et des avantages de linoculation pour la prévenir*, Hist. Acad. R. Sci. Paris (1760/1766), 1–45.

[3] N. F. Britton: *Essential Mathematical Biology*, Springer-Verlag, London, 2003.

[4] R. Dawkins: *The Selfish Gene*, Oxford Univ. Press, Oxford, 1989, český překlad: *Sobecký gen*, Mladá fronta, Praha, 1998.

[5] K. Dietz, J. A. P. Heesterbeck: *Daniel Bernoulli's epidemiological model revisited*, Math. Biosci. **180** (2002), 1–21.

[6] A. W. F. Edwards: *Foundations of Mathematical Genetics*, Cambridge Univ. Press, Cambridge, 2000.

[7] J. Flegr: *Zamrzlá evoluce aneb je to jinak, pane Darwin*, Academia, Praha, 2006.

[8] R. P. Feynman, R. B. Leighton, M. Sands: *The Feynman lectures on physics*, Addison-Wesley Publishing Company, 1966.

[9] R. A. Fisher: *On the dominance ratio*, Proc. R. Soc. Edinb. **42** (1922), 321–341.

[10] R. A. Fisher: *The Genetical Theory of Natural Selection*, Clarendon Press, Oxford, 1930.

[11] J. B. S. Haldane: *A mathematical theory of natural and artificial selection, Part V, Selection and mutation*, Proc. Camb. Philos. Soc. **23** (1927), 838–844.

[12] G. H. Hardy: *Mendelian proportions in a mixed population*, Science **28** (1908), 49–50.

[13] G. H. Hardy: *A Mathematician's Apology*, Cambridge Univ. Press, Cambridge, 1967, český překlad: *Obrana matematikova*, Prostor, Praha, 1999.

[14] J. Hofbauer, K. Sigmund: *Evolutionary Games and Population Dynamics*, Cambridge Univ. Press, Cambridge, 2002.

[15] J. Hofbauer, P. Schuster, K. Sigmund: *A note on evolutionarily stable strategies and game dynamics*, J. Theor. Biol. **81** (1979), 609–612.

[16] J. Kalas, M. Ráb: *Obyčejné diferenciální rovnice*, Masarykova univerzita, Brno, 2001.

[17] R. Kotecký: *Na rozhraní mezi matematikou a fyzikou*, v: P. Kůrka, A. Matoušek, B. Velický (ed.): *Spor o matematizaci světa*, Pavel Mervart, Červený Kostelec, 2011, 45–54.
[18] L. Kvasz: *Galileovská fyzika vo svetle Husserlovej fenomenologie*, Filosofický časopis **48** (2000), 373-399.
[19] J. G. Mendel: *Versuche űber Pflanzenhybriden*, Verh. Naturfosch. Ver. Brűnn **4** (1866), 3–47.
[20] P. Schuster, K. Sigmund: *Coyness, philandering and stable strategies*, Anim. Behavior **29** (1981), 186–192.
[21] J. Maynard Smith: *Evolution and the Theory of Games*, Cambridge Univ. Press, Cambridge, 1982.
[22] J. Maynard Smith, G. Price: *The logic of animal conflict*, Nature **246** (1973), 15–18.
[23] P. D. Taylor, L. Jonker: *Evolutionary stable strategies and game dynamics*, Math. Biosci **44** (1978), 145–156.
[24] W. Weinberg: *Űber den Nachweis der Verebung beim Menschen*, Jahresh. Wuertt. Ver. vaterl. Natkd. **64** (1908), 369-382.
[25] S. Wright: *Evolution in Mendelian populations*, Genetics **16** (1931), 97–159.

Zdeněk Pospíšil, Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita, Kotlářská 2, 611 37, Brno, Česká republika,
e-mail: pospisil@math.muni.cz

WEILOVO PÁROVÁNÍ NA ELIPTICKÝCH KŘIVKÁCH V KRYPTOSYSTÉMECH ZALOŽENÝCH NA TOTOŽNOSTI A KRYPTOGRAFICKY RANDOMIZOVANÉ ODPOVÍDACÍ TECHNIKY

MIROSLAV KUREŠ

Abstrakt. Článek shrnuje dvě autorovy přednášky: o kryptografii založené na totožnosti a o kryptograficky randomizovaných odpovídacích technikách. V obou případech se opíráme o eliptickou kryptografii, tedy systémy využívající eliptické křivky nad konečnými poli. Hlavní účelem článku je poskytnout přehledný text demonstrující, s jakými algebraickými problémy se lze v eliptické kryptografii setkat a také, v případě randomizovaných odpovídacích technik, i poněkud nečekané užití kryptografického protokolu v dotazníkových šetřeních.

Úvod

Text tohoto článku je členěn do dvou kapitol odpovídajících dvěma souvisejícím přednáškám autora: první se věnovala kryptografii založené na totožnosti, druhá kryptograficky randomizovaným odpovídacím technikám. Důraz je kladen na eliptickou kryptografii (ECC), již lze chápat i jako pojící bázi mezi tématy. Upření zájmu k ECC má své odůvodnění. Lze se totiž oprávněně domnívat, že nejpopulárnější algoritmus asymetrické kryptografie RSA je v současnosti také předmětem největšího úsilí o prolomení. ECC se opírá o řádově náročnější matematiku a přitom vykazuje i některé lepší vlastnosti (stejná bezpečnost s kratšími klíči).

Článek vyžaduje základní znalost kryptografických a některých algebraických pojmů a jistý vhled do problematiky.

1. Kryptosystémy založené na totožnosti a Weilovo párování na eliptických křivkách

1.1. Kryptosystémy s veřejným klíčem ve srovnání s kryptosystémy založenými na totožnosti

Připomeňme nejprve základní parametry kryptosystémů s veřejným klíčem ve srovnání s kryptosystémy založenými na totožnosti za účasti v kryptografii již oblíbených osob: Alice, Boba a Evy.

2010 *MSC.* Primární 94A60, 68W20; Sekundární 14H52.

Klíčová slova. Kryptografie založená na eliptických křivkách, Weilovo párování, randomizované odpovídací techniky.

Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

Kryptosystém s veřejným klíčem: je tvořen veřejným klíčem K_U a soukromým klíčem K_R. Klíč K_U je generován jednosměrnou funkcí z K_R. Pokud je Bob offline, nelze komunikovat. Pokud Eva (*(wo)man-in-the-middle*) přesvědčí Alici, že K_U je jiný, snadno pak dešifruje zprávy určené Bobovi. Tento *autentizační problém* (ověření identity) je řešen *certifikáty* poskytovanými *důvěryhodnou autoritou*.

Kryptosystém založený na identitě: veřejným klíčem K_U je jednoznačná identifikace Boba (např. telefonní číslo nebo e-mailová adresa). Certifikáty nejsou třeba. Je-li Bob offline, lze komunikovat (např. zpráva může čekat v *Short Message Service Center* a být odeslána později).

1.2. Eliptické křivky: základní pojmy

Dále uveďme základní pojmy kryptografie založené na eliptických křivkách.

Eliptické křivky. *Eliptickou křivkou* $\mathcal{E}$ nad polem $\mathbb{F}$ rozumíme algebraickou křivku třetího stupně s rovnicí

$$y^2 + a_1xy + a_3y = x^3 + a_2x^2 + a_4x + a_6,$$

kde $a_1, a_2, a_3, a_4, a_6 \in \mathbb{F}$ a kde tzv. *diskriminant* Δ eliptické křivky $\mathcal{E}$ je nenulový. Přitom

$$\begin{aligned}
\Delta &= -d_2^2d_8 - 8d_4^3 - 27d_6^2 + 9d_2d_4d_6\\
d_2 &= a_1^2 + 4a_2\\
d_4 &= 2a_4 + a_1a_3\\
d_6 &= a_3^2 + 4a_6\\
d_8 &= a_1^2a_6 + 4a_2a_6 - a_1a_3a_4 + a_2a_3^2 + a_4^2.
\end{aligned}$$

Je-li $\mathbb{F}_q = \mathbb{F}_{p^n}$ (p prvočíslo, $n \in \mathbb{N}$ konečné pole s charakteristikou různou od 2 a 3), pak vhodnou změnou souřadnic lze Weierstrassovu rovnici transformovat na rovnici

$$y^2 = x^3 + ax + b.$$

Je-li $\mathbb{F}_q$ konečné pole s charakteristikou 2, pak vhodnou změnou souřadnic lze Weierstrassovu rovnici transformovat na rovnici

$$y^2 + xy = x^3 + ax^2 + b$$

(tzv. *nesupersingulární* eliptická křivka) nebo na rovnici

$$y^2 + cy = x^3 + ax + b$$

(tzv. *supersingulární* eliptická křivka). Obdobná věta platí i pro pole charakteristiky 3.

Bodem eliptické křivky $\mathcal{E}$ rozumíme každý bod $[x, y]$ se souřadnicemi $x, y \in \mathbb{F}$ splňujícími její rovnici a dále bod ∞. Nad body eliptické křivky (nadále již bereme $\infty \in \mathcal{E}$) lze zavést binární operaci značenou $+$ a nazvanou *sečno-tečnové sčítání* takto: jsou-li dva body $P = [x_1, y_1]$, $Q = [x_2, y_2]$ eliptické křivky $\mathcal{E}$ různé, vedeme skrze ně přímku; ta protne $\mathcal{E}$ ještě v dalším bodě a $R = P + Q$ vezmeme jako bod osově souměrný s tímto třetím bodem podle osy x. (Je-li přímka rovnoběžná s osou y, třetí bod $\mathcal{E}$ na ní již neexistuje, pak klademe $P + Q = \infty$.) Dále, pro součet $P+P$ vedeme bodem P tečnu; ta protne $\mathcal{E}$ ještě v dalším bodě a $R = P+P$ vezmeme jako bod osově souměrný s tímto bodem podle osy x. (Opět, je-li přímka

rovnoběžná s osou y, klademe $P + P = \infty$.) Součet libovolného bodu P eliptické křivky $\mathcal{E}$ s bodem ∞ položíme roven P.

Vzorce pro sečno-tečnové sčítání (pro pole s p různým od 2 i od 3):
(případ $R = P + Q; R = [r_1, r_2], P = [p_1, p_2], Q = [q_1, q_2]$)

$$r_1 = \left(\frac{q_2 - p_2}{q_1 - p_1}\right)^2 - p_1 - q_1 \qquad r_2 = \left(\frac{q_2 - p_2}{q_1 - p_1}\right)(p_1 - r_1) - p_2$$

(případ $R = 2P = P + P; R = [r_1, r_2], P = [p_1, p_2]$)

$$r_1 = \left(\frac{3p_1^2 + a}{2p_2}\right)^2 - 2p_1 \qquad r_2 = \left(\frac{3p_1^2 + a}{2p_2}\right)(p_1 - r_1) - p_2$$

1.3. Eliptické křivky: vlastnosti grupy $(\mathcal{E}, +)$

Nyní je $(\mathcal{E}, +)$ grupa. *Řádem eliptické křivky* $\mathcal{E}$ pak rozumíme řád této grupy čili počet bodů $\mathcal{E}$; značíme ho $\#\mathcal{E}$. Je-li $\mathcal{E}$ eliptická křivka nad konečným polem $\mathbb{F}_q$, pro její řád $\#\mathcal{E}(\mathbb{F}_q)$ platí odhad

$$q + 1 - 2\sqrt{q} \leq \#\mathcal{E}(\mathbb{F}_q) \leq q + 1 + 2\sqrt{q}$$

(*Hasseho interval*). Existuje tedy $t \in \mathbb{Z}$ takové, že

$$\#\mathcal{E}(\mathbb{F}_q) = q + 1 - t, \quad \text{kde } |t| \leq 2\sqrt{q}.$$

Řád eliptické křivky lze pomocí t a Legendrova symbolu určit takto:

$$t = -\sum_{x \in \mathbb{F}_p} \left(\frac{x^3 + ax + b}{p}\right)$$

Přímé užití tohoto vztahu se nazývá *naivní algoritmus* pro určení řádu eliptické křivky.

1.3.1. Torzní podgrupy. Pro $m \in \mathbb{N}$ bod P křivky $\mathcal{E}$ splňující $mP = \infty$ nazveme *m-tý torzní bod*. Množina všech m-tých torzních bodů se značí $\mathcal{E}[m]$, se zavedeným sčítáním bodů je $\mathcal{E}[m]$ podgrupou grupy $\mathcal{E}$, nazýváme ji *m-tá torzní grupa*.

Je-li m libovolným násobkem řádu bodu P, pak $P \in \mathcal{E}[m]$. První torzní podgrupa je evidentně triviální (obsahuje pouze bod ∞), zatímco $\#\mathcal{E}$-tá torzní podgrupa už je rovna $\mathcal{E}$. (Nic nového nepřináší, uvažujeme-li $m > \#\mathcal{E}$.) Bod ∞ je prvkem všech m-tých torzních grup a dále je zřejmé, že v případě nesoudělného m a $\#\mathcal{E}$ je také m-tá torzní podgrupa vždy triviální.

Jak vypadá druhá torzní grupa eliptických křivek? Pro její body platí $P + P = \infty$, což je možné jen pro nulovou y-ovou souřadnici bodu P. Tedy jde o to, zda existuje kořen rovnice $x^3 + ax + b = 0$ (v $\mathbb{F}_p$). Pokud ano, je řád $\#\mathcal{E}(\mathbb{F}_p)$ sudý, a tedy t je sudé, neexistuje-li kořen $x^3 + ax + b = 0$ (v $\mathbb{F}_p$), je t liché.

Obecně nejsou torzní podgrupy grupy G cyklické, neboť sama grupa G nemusí být cyklická; nejsou cyklické ani pro případ m, které je menší než řád grupy G: například druhá torzní grupa grupy $G = \mathbb{Z}_4 \oplus \mathbb{Z}_6$ je tvořena body $(0,0)$, $(2,0)$, $(0,3)$, $(2,3)$ a evidentně není generovaná jediným prvkem.

Pokud jde o grupu $\mathcal{E}$ nad polem $\mathbb{F}_q$, pak ta je vždy izomorfní grupě $\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2}$, přičemž n_2 dělí jak n_1, tak $q - 1$ (Teorém 3.12, [6]). Pak ovšem $\#\mathcal{E} = n_1 n_2$. Je-li $n_2 = 1$, je grupa $\mathcal{E}$ cyklická. Je-li $n_2 > 1$ malé přirozené číslo $(2, 3, 4, \dots)$, říkáme,

že grupa $\mathcal{E}$ je *téměř cyklická*. V cyklické grupě existuje bod, jehož řád je roven již přímo $\#\mathcal{E}$. Pro určení řádu eliptické křivky je ale také výhodné, je-li téměř cyklická, neboť v ní existují body dostatečně vysokého řádu, jejichž násobek patří do Hasseho intervalu.

1.4. Eliptická kryptografie s veřejným klíčem

Uvedeme nyní, v čem spočívá podstata eliptické kryptografie s veřejným klíčem (ECC). Pro jednoduchost vezmeme prvočíselné polem $\mathbb{F}_p$, nad kterým uvažujeme eliptickou křivku $\mathcal{E}$ a její bod $P \in \mathcal{E}$. *Řádem bodu* P rozumíme nejmenší $n \in \mathbb{N}$ takové, že $nP = \infty$. Řád každého bodu P eliptické křivky $\mathcal{E}$ dělí řád této křivky. (To je známý Lagrangeův teorém z teorie grup.) Bod P vyberme tak, aby jeho řádem bylo prvočíslo n. Dále vyberme nějaké $k \in [1, n-1]$ a spočtěme $Q = kP$. Údaje o poli a eliptické křivce jsou tzv. *definiční parametry* a pokládají se za známé. Veřejným klíčem jsou body P a Q a soukromým klíčem číslo k.

Algoritmus. Základní ElGamal šifrování pomocí eliptických křivek.

Vstup: Definiční parametry, veřejný klíč P, Q, zpráva m.

Výstup: Šifrovaná zpráva (C_1, C_2).

1. Vyjádři m jako bod M eliptické křivky.
2. Vyber $a \in [1, n-1]$.
3. Spočti $C_1 = aP$.
4. Spočti $C_2 = M + aQ$.
5. Vrať (C_1, C_2).

Algoritmus. Základní ElGamal dešifrování pomocí eliptických křivek.

Vstup: Definiční parametry, veřejný klíč P, Q, soukromý klíč k, šifrovaná zpráva (C_1, C_2).

Výstup: Zpráva m.

1. Spočti $M = C_2 - kC_1$.
2. Převeď M na m.
3. Vrať m.

1.5. Bilineární zobrazení grup a Weilovo párování

Uvažujme dvě grupy $G = (G, +)$, $H = (H, \cdot)$, $k \in \mathbb{Z}$, a zobrazení $\phi\colon G \times G \to H$ splňující

$$\begin{aligned} \phi(g_1 + g_2, g_3) &= \phi(g_1, g_3) \cdot \phi(g_2, g_3) \\ \phi(kg_1, g_2) &= (\phi(g_1, g_2))^k \\ \phi(g_1, g_2 + g_3) &= \phi(g_1, g_2) \cdot \phi(g_1, g_3) \\ \phi(g_1, kg_2) &= (\phi(g_1, g_2))^k . \end{aligned}$$

Takové zobrazení nazveme *bilineární zobrazení*. (Užíváme zde aditivní notaci pro G a multiplikativní notaci pro H.)

V případě eliptických křivek můžeme uvažovat např. $G = (\mathcal{E}, +)$ a $H = (\mathbb{F}_q - \{0\}, \cdot)$, pro dále uvedené Weilova párování je volba grup speciálnější.

Weilovo párování je zobrazení

$$e\colon \mathcal{E}[m] \times \mathcal{E}[m] \to U_m,$$

kde $U_m = (U_m, \cdot)$ je grupa m-tých odmocnin jedničky v $\bar{\mathbb{F}}_p$ ($\bar{\mathbb{F}}_p$ je algebraický uzávěr $\mathbb{F}_q = \mathbb{F}_{p^n}$).

1.6. Weilovo párování v kryptografii založené na totožnosti

Důvěryhodná autorita je označována jako PKG (*private key generator*). Ta zvolí univerzální tajný klíč s. Poté zveřejní: rovnici eliptické křivky, její základní bod P, veřejný klíč systému sP a hašovací funkci h. Každý uživatel má veřejný klíč $K_U = Q_{\mathrm{ID}}$ (bod eliptické křivky vycházející z identity) a soukromý klíč $K_R = sQ_{\mathrm{ID}}$, který obdrží od PKG. Odesílatel zprávy M vybere náhodně číslo r a posílá

$$(U, V) = (rP, M + \mathsf{h}(e(Q_{\mathrm{ID}}, sP)^r))\,.$$

Adresát pak za použití svého soukromého klíče sQ_{ID} z (U, V) spočte

$$M = V + \mathsf{h}(e(sQ_{\mathrm{ID}}, U)).$$

1.6.1. Závěrečné poznámky k Weilovu párování. Kryptografie založená na identitě je vhodná zejména pro šifrování mobilními telefony. Mezi eliptické křivky doporučené standardy patří

$$y^2 + xy = x^3 + x^2 + b$$

nad binárními poli $\mathbb{F}_{2^{163}}$, $\mathbb{F}_{2^{233}}$, $\mathbb{F}_{2^{409}}$ (síla šifrování roste).

Např. pro $q = 2^{409}$ musí n_2 dělit jak n_1, tak některé z čísel

$$4480666067023,$$

$$76025626689833,$$

$$3881196575913244673719425770871246487895686937951690944453838587$$
$$6764072695131586617955811936945129,$$

tzn. pro drtivou většinu křivek (cvičení: nebo pro všechny?) nad tímto polem je grupa eliptické křivky cyklická.

Bezpečnost kryptosystému založeného na identitě pak plyne z obtížnosti tzv. *Diffieho-Hellmanova problému pro cyklické grupy.*

2. Randomizované odpovídací techniky a kryptograficky randomizované odpovídací techniky

2.1. Motivace pro nepřímé otázky a randomizaci

Kryptografie má zajímavé použití i v dotazníkových šetřeních. Uplatní se při řešení problému dotazování na záležitosti, u nichž lze očekávat, že respondent bude mít z nejrůznějších důvodů obavu odpovědět podle pravdy. Základní uvažovaná otázka je tato: *Patříte do („stigmatizované") skupiny A?*

Z literatury věnující se problematice uveďme např. knihu [2], která je přehlednou monografií o randomizovaných technikách a technikách nepřímého dotazování.

2.2. Warnerova RRT

Klasická RRT (Warner 1965) vychází z následujícího schématu:

p_{ct} (známá) pravděpodobnost, že respondent odpoví na otázku pravdivě ($> \frac{1}{2}$, resp. stačí $\neq \frac{1}{2}$)

π_A skutečný podíl populace patřící do A

p_{yes} podíl populace s odpovědí ano pak

$$p_{\mathsf{yes}} = p_{\mathsf{ct}}\pi_A + (1 - p_{\mathsf{ct}})(1 - \pi_A);$$

pro N respondentů a L odpovědí ano je bodový odhad $\widehat{p_{\mathsf{yes}}}$ hodnoty p_{yes} roven $\widehat{p_{\mathsf{yes}}} = \frac{L}{N}$, tzn. můžeme spočíst bodový odhad $\widehat{\pi_A}$ hodnoty π_A jako

$$\widehat{\pi_A} = \frac{\widehat{p_{\mathsf{yes}}} - (1 - p_{\mathsf{ct}})}{2p_{\mathsf{ct}} - 1} = \frac{p_{\mathsf{ct}} - 1}{2p_{\mathsf{ct}} - 1} + \frac{L}{N}\frac{1}{2p_{\mathsf{ct}} - 1}.$$

$$\mathrm{E}\,\widehat{\pi_A} = \cdots = \pi_A$$

$$\mathrm{Var}\,\widehat{\pi_A} = \cdots = \frac{1}{N}\left(\frac{1}{16(p_{\mathsf{ct}} - \frac{1}{2})^2} - \left(\pi_A - \frac{1}{2}\right)^2\right)$$

Předpokládejme, že například víme, že $\frac{5}{6}$ populace odpovídá na každou otázku pravdivě a $\frac{1}{6}$ vždy lže.

Potom $p_{\mathsf{yes}} \in \left[\frac{1}{6}, \frac{5}{6}\right]$. Bude-li pak $\widehat{p_{\mathsf{yes}}} = \frac{2}{3}$, je

$$\widehat{\pi_A} = \frac{\frac{2}{3} - (1 - \frac{5}{6})}{2 \cdot \frac{5}{6} - 1} = \frac{3}{4}$$

Protože obvykle nevíme, jaký podíl populace na otázku odpoví pravdivě a jaký podíl zalže, náhodně respondentovi vygenerujeme nebo si vygeneruje sám (*randomizace*), zda odpoví na otázku nebo na její negaci. Skutečnost, zda respondent odpovídá na otázku nebo na její negaci zůstane pro tazatele utajená, respondent o tomto utajení ví, a nemá proto důvod lhát.

Příklad:

Otázka příslušnosti k A (otázka 1): Přiznáváte, že jste se někdy dopustil daňového podvodu?

Negace (otázka 2): Můžete říct, že jste se nikdy nedopustil daňového podvodu?

Hoďte si kostkou. Hodíte-li šestku, odpovězte na otázku 2. Hodíte-li něco jiného, odpovězte na otázku 1.

Dostaneme-li $\frac{2}{3}$ odpovědí ano, máme

$$\widehat{\pi_A} = \frac{\frac{2}{3} - (1 - \frac{5}{6})}{2 \cdot \frac{5}{6} - 1} = \frac{3}{4}$$

2.3. Kryptograficky randomizovaná (Warnerova) odpovídací technika

Randomizace odpovědi respondenta nemusí být ponechána na respondentovi: může ji provést důvěryhodná autorita.

CRRT protokol ideálního světa:

$\mathcal{I}$ tazatel

$\mathcal{R}$ respondent

$\mathcal{T}$ důvěryhodná autorita

$$\mathcal{R} \xrightarrow{t_{\mathcal{R}}} \boxed{\Phi_{p_{\mathsf{ct}}}(t_{\mathcal{R}}) = r_{\mathcal{R}}} \xrightarrow{r_{\mathcal{R}}} \mathcal{I}$$

(kde $\Phi_{p_{\mathsf{ct}}}(x)$ je randomizující funkce, jejímž výstupem je x s pravděpodobností p_{ct})

V reálném světě je důvěryhodná autorita nahrazena kryptografickým protokolem. Pro ten se definují tři vlastnosti — požadavky:

- PR *soukromí respondenta*
- PI *soukromí tazatele*
- C *korektnost*
- *slabý CRRT protokol*: PR, C
- *silný CRRT protokol*: PR, PI, C

Nechť p je velké prvočíslo a q, $q|(p-1)$, jiné prvočíslo. Pak má $\mathbb{Z}_p$ jedinou multiplikativní podgrupu G řádu q. Nechť g a h jsou dva z generátorů G (jejich vzájemné diskrétní logaritmy nejsou známy). Soukromý parametr: $k = q$, veřejný klíč $K = (g, h)$.

Zpráva $\mu \in \mathbb{Z}_q$ je podepsána náhodným $\rho \in \mathbb{Z}_q$ pomocí funkce

$$\mathsf{C}_K(\mu, \rho) = g^{\mu} h^{\rho}.$$

$$p_{\mathsf{ct}} = \frac{l}{n}$$

$$d = \left\lceil \frac{1}{1 - p_{\mathsf{ct}}} \right\rceil$$

Idea protokolu spočívá ve skutečnosti, že alespoň jedno z čísel $\mu + \nu + il \mod n$ musí ležet v intervalu $[0, d-1]$ a alespoň jedno z těchto čísel musí ležet v intervalu $[l, n-1]$.

Přípravný krok:

- $\mathcal{R}$ vybere náhodně μ z intervalu $[0, n-1]$
- $\mathcal{I}$ vybere náhodně ν z intervalu $[0, n-1]$ a σ z intervalu $[0, d-1]$

Interaktivní krok:

- $\mathcal{R}$ spočte $\mathsf{C}_K(t_{\mathcal{R}}, \mu)$ a odešle výsledek $\mathcal{I}$
- $\mathcal{I}$ spočte $y = \mathsf{C}_K(\sigma, \rho)$ pro náhodně vybrané ρ; ν a y odešle $\mathcal{R}$ spolu s informací že y je C_K-funkcí nějakého $i \in [0, d-1]$
- $\mathcal{R}$ verifikuje informaci; spočte hodnoty μ'_i tak, že $\mu'_i = t_{\mathcal{R}} \iff (\mu + \nu + il \mod n) < l$; podepíše σ a pošle podpis spolu s $\{\mu'_i\}$ pro všechna $i \in [0, d-1]$: $[(\mu + \nu + il \mod n) < l]$
- Potom $\mathcal{I}$ položí $r_{\mathcal{R}} = \mu'_\sigma$; to doplní podpisem $\mathcal{R}$, který může být ověřen $\mathcal{R}$ i třetími stranami

2.4. Modifikované a zobecněné metody

2.4.1. Greenbergova *et al.* **metoda.** V Greenbergově metodě je negace otázky nahrazena jinou otázkou, často nesouvisející s předmětem výzkumu, *neškodnou*. Na tuto otázku se předpokládá pravdivá odpověď, pravděpodobnost odpovědi ano je známá, v nejjednodušší variantě je pak 1 nebo 0.

Příklad:

Otázka příslušnosti k A (otázka 1): Přiznáváte, že jste se někdy dopustil daňového podvodu?

Neškodná otázka (otázka 2): Padl vám při hodu mincí orel?

Hoďte mincí. Padne-li vám orel, odpovězte na otázku 2. Padne-li vám hlava, odpovězte na otázku 1.

Dostaneme-li $\frac{2}{3}$ odpovědí ano, máme

$$\widehat{\pi_A} = \frac{1}{2} \cdot \frac{2}{3} + \frac{1}{2} \cdot 1 = \frac{5}{6}.$$

2.4.2. Polychotomická RRT. V klasické RRT se zkoumá příslušnost ke skupině A: buď ano nebo ne (*dichotomická RRT*).

V *polychotomické RRT* může být odpovědí více, přičemž míra „stigmatizace" může být různá.

Příklad.

Daňového podvodu jste se dopustil:

- zcela vědomě a úmyslně, pro vlastní obohacení
- z nedbalosti, pro zjednodušení administrativy, kvůli špatné evidenci dokladů, apod.
- nedopustil jsem se daňového podvodu

2.4.3. Metoda tří karet. Vyvinuta pro United States Government Accountability Office (GAO) pro zjištění počtu ilegálně usazeného (španělsky hovořícího) obyvatelstva. Respondenti určí svůj status postupně na třech kartách 1, 2 a 3, které vhazují do boxů A, B a C. Citlivý status není nikdy v samostatné skupině, dochází ale k přeskupování.

2.4.4. Závěrečné poznámky k použití různých kryptosystémů v CRRT. Navržený protokol je připraven (po případných modifikacích) k implementaci v různých kryptosystémech s veřejným klíčem. Připomeňme například, že v mobilní telefonii jsou užity kryptosystémy založené na eliptických křivkách (kvůli výrazně menší hardwarové náročnosti, než má nejrozšířenější RSA). Nabízí se zkusit zobecnit Pedersenovo schéma (pro $\mathbb{Z}_p$) pro $\mathbb{Z}_{n_1} \oplus \mathbb{Z}_{n_2}$.

Reference

[1] A. Ambainis, M. Jakobsson, H. Lipmaa: *Cryptographic randomized response techniques*, v: Public Key Cryptography – PKC 2004, Vol. 2947 of Lecture Notes in Computer Science, 2004, 425–438.

[2] A. Chaudhuri: *Randomized Response and Indirect Questioning Techniques in Surveys*, Chapman & Hall (CRC), Taylor & Francis, 2011.

[3] J. A. Droitcour, E. M. Larson, F. J. Scheuren: *The three card method: Estimating sensitive survey itemswith permanent anonymity of response*, v: Proceedings of the American Statistical Association, Social Statistics Section, 2001.

[4] B. G. Greenberg, A. A. Abul-Ela, W. R. Simmons, D. G. Horvitz: *The unrelated question randomized response model: Theoretical framework*, J. Amer. Statist. Assoc. **64** (1969), 520–539.

[5] J. S. Hwu, R.J. Chen, Y.-B. Lin: *An efficient identity-based cryptosystem for end-to-end mobile security*, IEEE Transactions **5** (2006), 2586–2593.

[6] D. Hankerson, A. Menezes, S. Vanstone: *Guide to Elliptic Curve Cryptography*, Springer, 2004.

[7] R. Sakai, M. Kasahara: *ID based cryptosystems with pairing on elliptic curve*, Cryptology ePrint Archive, Report 2003/054.
[8] S. L. Warner: *Randomized response: A survey technique for eliminating evasive answer bias*, J. Amer. Statist. Assoc. **60** (1965), 63–69.

Miroslav Kureš, Ústav matematiky, Fakulta strojního inženýrství, Vysoké učení technické v Brně, Technická 2, 616 69 Brno, Česká republika,
e-mail: kures@fme.vutbr.cz

OD KOMPOZITNÍCH MATERIÁLŮ KE SLABÉ KONVERGENCI

JAN FRANCŮ

Abstrakt. Matematické modelování kompozitních materiálů využívá tzv. homogenizaci, při které heterogenní materiál s jemnou periodickou strukturou nahradíme homogenním materiálem, který má z makroskopického hlediska stejné vlastnosti. Matematický přístup spočívá ve studiu posloupnosti řešení parciálních diferenciálních rovnic s periodickými koeficienty se zmenšující se periodou. Homogenizace umožňuje počítat makroskopické vlastnosti materiálu z vlastností jednotlivých složek a jejich geometrického uspořádání.

Koeficienty se zmenšující se periodou studované při homogenizaci nekonvergují (ani bodově ani v normě), konvergují však slabě. Článek je zaměřen na tuto slabou konvergenci a její vlastnosti. Příjemnou vlastností je kompaktnost: každá omezená posloupnost funkcí obsahuje slabě konvergentní podposloupnost. Nepříjemnou vlastností je však nemožnost přechodu k limitě způsobené ztrátou informace ve slabé limitě. Pokračování tohoto článku se bude zabývat řešením tohoto problému.

1. Úvod

Kompozitní materiály hrají velmi důležitou roli v mnoha technických oborech. Kombinace látek různých vlastností umožňuje získat nové materiály výjimečných vlastností. Například ve sklolaminátu jsou pevná křehká skleněná vlákna spojena pryskyřicí do houževnatého materiálu, podobně v železobetonu je beton vyztužen ocelovými pruty, které podstatně zvyšují pevnost betonu v tahu.

Matematické modelování, tj. výpočet chování těchto materiálů, přináší základní problém: pokud bychom vycházeli z vlastností jednotlivých složek a chtěli pro výpočet popsat jemnou strukturu materiálu – tisíce tenkých vláken – museli bychom volit ještě jemnější triangulaci výpočetní oblasti, což by vedlo na milióny rovnic pro milióny neznámých. Jiný přístup spočívá ve tvorbě speciálních modelů kompozitního materiálu s řadou dalších parametrů, které je nutno měřit.

Ukazuje se však, že stačí zkoumaný heterogenní materiál nahradit myšleným „ekvivalentním“ homogenním materiálem, který je popsán konstantními koeficienty. Vzniká přitom otázka, zda tento tzv. homogenizovaný materiál lze popsat rovnicemi stejného typu s konstantními koeficienty. Pokud ano, otázkou je, jak spočítat tyto tzv. homogenizované koeficienty, abychom nemuseli vlastnosti materiálu měřit na vzorku, který by se musel vyrobit.

2010 *MSC.* Primární 35B27.

Klíčová slova. Kompozitní materiály, homogenizace, slabá konvergence.

Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

Matematický přístup k tomuto problému navržený Ivo Babuškou [1] spočívá v tom, že místo jednoho materiálu s periodickou strukturou uvažujeme posloupnost materiálů se zjemňující se strukturou. V matematické formulaci studujeme chování řešení posloupnosti parciálních diferenciálních rovnic s periodickými koeficienty se zmenšující se periodou. Tyto koeficienty však nekonvergují v obvyklém smyslu, ale jen tzv. slabě, viz [7].

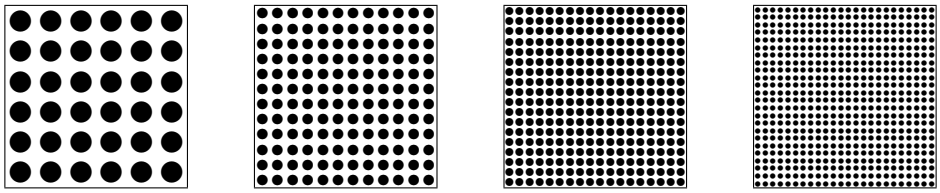
Slabá konvergence funkcí v normovaných prostorech integrovatelných funkcí má řadu vlastností, kterými se liší od obvyklé, tzv. silné konvergence. V příspěvku se budeme věnovat hlavně slabé konvergenci funkcí. Po definici slabé konvergence v normovaných prostorech si ukážeme případy slabé konvergence v konkrétních prostorech. Začneme prostory konečné dimenze, ve kterých slabá a silná konvergence splývají. V prostorech nekonečných posloupností a integrovatelných funkcí se již obě konvergence liší, což si ukážeme na příkladech.

Zatímco v prostorech konečné dimenze každá omezená uzavřená množina je kompaktní, zejména každá omezená posloupnost obsahuje konvergentní podposloupnost, v nekonečně rozměrných prostorech již toto neplatí: uvedeme příklad posloupnosti, ze které nelze vybrat konvergentní podposloupnost. V reflexivních prostorech (viz [7]) však každá omezená posloupnost obsahuje podposloupnost slabě konvergentní, což je velmi důležité v řadě aplikací. Tato „příjemná“ vlastnost slabé konvergence je „vykoupena“ problémy při limitních přechodech: například limita součinu dvou slabě konvergentních posloupností se nerovná součinu limit těchto posloupností. Řešením těchto problémů se bude zabývat pokračování tohoto článku.

Ačkoliv praktický význam mají zejména funkce na dvojrozměrné a trojrozměrné oblasti, teorie je budována pro N-rozměrný případ. My se budeme zabývat hlavně jednorozměrnými příklady, protože většina jevů se projeví již zde. Seznam literatury k tomuto tématu má tisíce položek, uvedeme proto jen několik článků v češtině a monografie [3], [2], [4], ve kterých lze najít odkazy na další literaturu.

2. Homogenizace

Homogenizací rozumíme postup, při kterém heterogenní materiál s periodickou strukturou nahradíme myšleným „ekvivalentním“ materiálem homogenním, který má na makroskopické úrovni stejné vlastnosti. V matematickém pojetí diferenciální rovnici s periodickými koeficienty nahradíme rovnicí s konstantními koeficienty dávající globálně stejné řešení.



Obr. 1. Příklad posloupnosti materiálů se zjemňující se strukturou.

Matematický přístup navržený Babuškou v roce 1973, viz [1], spočívá v tom, že místo jednoho materiálu s periodickou strukturou uvažujeme posloupnost materiálů se zjemňující se strukturou, v matematické formulaci uvažujeme posloupnost rovnic určených posloupností periodických koeficientů se zmenšující se periodou.

Než přistoupíme k formulaci problému homogenizace, zavedeme základní pojmy. Praktický význam má homogenizace rovnic na oblastech Ω v prostoru dimenze $N = 2$ a $N = 3$, teorie homogenizace však platí i pro $N > 3$. Body (prvky) prostoru $\mathbb{R}^N$ budeme označovat $x \equiv [x_1, \dots, x_N]$, $y \equiv [y_1, \dots, y_N]$, atd.

2.1. Periodické funkce se zmenšující se periodou

Pro jednoduchost základní periodou Y bude v případě $N = 2$ jednotkový čtverec a v případě $N = 3$ jednotková krychle, obecně $Y = \langle 0, 1) \times \cdots \times \langle 0, 1)$.

Funkci $a(y)$ definovanou na $\mathbb{R}^N$ nazveme *Y-periodickou*, jestliže je periodická s periodou 1 v každé proměnné, tj.

$$a(y_1 + k_1, \dots, y_N + k_N) = a(y_1, \dots, y_N)\,, \quad \forall y \in \mathbb{R}^N,\ \forall k \in \mathbb{Z}^N.$$

Pokud funkce a má další proměnné, např. x, budeme říkat, že funkce $a(x, y)$ je Y-periodická v y.

Budeme se zabývat posloupností funkcí se zmenšující se periodou ε. Místo přirozených čísel budeme posloupnosti „číslovat“ klesající posloupností malých čísel $\varepsilon_1, \varepsilon_2, \varepsilon_3, \dots$ konvergujících k nule: tj. $a^{\varepsilon_1}, a^{\varepsilon_2}, a^{\varepsilon_3}, \dots$, pro lepší čitelnost budeme psát jenom a^ε. Index ε má zde význam velikosti periody εY.

Buď Ω omezená oblast v $\mathbb{R}^N$ s „rozumnou“ hranicí. Pro Y-periodickou funkci $a(y)$ a $\varepsilon > 0$ vztah

$$a^\varepsilon(x) = a\left(\frac{x}{\varepsilon}\right) \equiv a\left(\frac{x_1}{\varepsilon}, \dots, \frac{x_N}{\varepsilon}\right), \quad x \in \Omega\,, \tag{2.1}$$

definuje εY-periodickou funkci na Ω. Pro posloupnost $\{\varepsilon\}$ klesající k nule tak dostáváme posloupnost funkcí $a^\varepsilon(x)$ se zmenšující se periodou.

2.2. Modelová úloha

Uvažujme lineární eliptickou parciální diferenciální rovnici druhého řádu

$$-\operatorname{div}(a(x)\,\nabla u) \equiv -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left(a(x)\frac{\partial u}{\partial x_i}\right) = f\,, \qquad x \in \Omega\,, \tag{2.2}$$

s vhodnou okrajovou podmínkou na hranici $\Gamma = \partial\Omega$, například $u|_\Gamma = 0$. Pokud koeficient $a(x)$ není spojitý, v rovnici „vnější“ derivaci musíme brát v zobecněném smyslu. Poznamenejme, že pro omezenou po částech spojitou funkci $a(x) \geq \alpha > 0$ a $f \in L^2(\Omega)$ uvedená okrajová úloha má právě jedno (tzv. slabé) řešení, viz [9].

Rovnici lze fyzikálně interpretovat jako rovnici ustáleného vedení tepla v desce tvaru Ω ($N = 2$) nebo v tělese ($N = 3$) zabírajícím objem Ω. Neznámá $u(x)$ je teplota, koeficient $a(x)$ popisuje vlastnosti izotropního materiálu (poměr tepelné vodivosti a objemového měrného tepla) a $f(x)$ hustota výkonu vnitřních zdrojů tepla. Okrajová podmínka $u = 0$ říká, že na okraji Γ je udržovaná nulová teplota. Rovnice popisuje i difuzi a kroucení tyče, viz např. [8].

2.3. Homogenizace – formulace úlohy

Při homogenizaci zkoumáme posloupnost okrajových úloh pro rovnice s koeficienty se zmenšující se periodou. Buď $a(y)$ navíc Y-periodická funkce. Potom pro každé $\varepsilon > 0$ máme εY-periodický koeficient $a^\varepsilon(x) = a(\frac{x}{\varepsilon})$ a tím i rovnici

$$-\operatorname{div}(a^\varepsilon(x)\,\nabla u^\varepsilon) \equiv -\sum_{i=1}^{N} \frac{\partial}{\partial x_i}\left(a^\varepsilon(x)\,\frac{\partial u^\varepsilon}{\partial x_i}\right) = f\,, \qquad x \in \Omega\,, \tag{2.3}$$

která s okrajovou podmínkou $u^\varepsilon|_\Gamma = 0$ tvoří okrajovou úlohu pro řešení u^ε.
V tzv. slabé formulaci na prostoru funkcí $H_0^1(\Omega)$[1] tato úloha zní:

Hledáme funkci $u^\varepsilon \in H_0^1(\Omega)$ *splňující integrální identitu*

$$\int_\Omega \sum_{i=1}^{N} a^\varepsilon(x) \frac{\partial u^\varepsilon}{\partial x_i}\frac{\partial v}{\partial x_i}\,\mathrm{d}x = \int_\Omega f v\,\mathrm{d}x\,, \qquad \forall v \in H_0^1(\Omega). \tag{2.4}$$

Za výše uvedených předpokladů pro každé $\varepsilon > 0$ úloha má právě jedno řešení. Pro posloupnost $\{\varepsilon\}$ parametrů $\varepsilon_n = \frac{1}{n}$ tak máme posloupnost úloh (2.3), a tím odpovídající posloupnost řešení u^ε. Homogenizace se zabývá studiem chování této posloupnosti řešení.

2.4. Problémy, kterými se zabývá homogenizace

Při homogenizaci řešíme následující otázky:

1. Konverguje posloupnost řešení u^ε? Pokud ano, v jakém smyslu?
2. Pokud u^ε konverguje k nějaké funkci u^*, je limita u^* řešením okrajové úlohy stejného typu, ale s konstantními tzv. homogenizovanými koeficienty?
3. Jak lze spočítat homogenizované koeficienty?
4. Jak lze zlepšit aproximaci řešení u^ε pro konkrétní $\varepsilon > 0$ pomocí u^*?

Pro představu na uvedené otázky uvedeme jen stručné odpovědi bez důkazů. V případě homogenizace úlohy (2.3) lze odvodit, viz [5], [6], následující výsledky:

1. Posloupnost řešení u^ε pro $\varepsilon \to 0$ konverguje k funkci, kterou označíme u^*.
2. Limita u^* je řešením opět eliptické rovnice druhého řádu

$$-\operatorname{div}(b\,\nabla u^*) \equiv -\sum_{i,j=1}^{N} b_{ij}\,\frac{\partial^2 u^\varepsilon}{\partial x_i\,\partial x_j} = f,$$

ale místo jednoho koeficientu $a(y)$ máme matici koeficientů $\{b_{ij}\}_{i,j=1}^N$. I když původní složky materiálu jsou izotropní, homogenizovaný materiál může být anizotropní.

---

[1]Symbol $H_0^1(\Omega)$ značí Sobolevův prostor, tj. prostor funkcí $u(x)$ na Ω, které mají nulové hodnoty na hranici $\partial\Omega$ a integrovatelné druhé mocniny parciálních derivací, viz např. [9]. Prostor $H_0^1(\Omega)$ je definován jako zúplnění prostoru hladkých funkcí nulových na $\partial\Omega$ v normě

$$\|u\| = \left[\int_\Omega \left(|u(x)|^2 + \left|\frac{\partial u}{\partial x_1}(x)\right|^2 + \cdots + \left|\frac{\partial u}{\partial x_N}(x)\right|^2\right)\mathrm{d}x\right]^{1/2}.$$

3. Homogenizované koeficienty lze spočítat z vlastností jednotlivých složek a jejich geometrického rozložení v periodě Y. Lze odvodit, viz [5], [6], že

$$b_{ij} = \int_Y \left(a(y)\,\delta_{ij} - a(y)\,\frac{\partial \chi^j}{\partial y_i}(y) \right) \mathrm{d}y\,,$$

kde δ_{ij} je tzv. Kroneckerovo delta: $\delta_{ij} = 1$ pro $i = j$ a $\delta_{ij} = 0$ pro $i \neq j$ a pomocné funkce $\chi^j(y)$ jsou Y-periodická slabá řešení rovnice

$$-\operatorname{div}\bigl(a\,\nabla\chi^j\bigr) \equiv -\sum_{i=1}^{N} \frac{\partial}{\partial y_i}\left(a(y)\,\frac{\partial \chi^j}{\partial y_i}(y) \right) = -\frac{\partial a}{\partial y_j}(y)\,, \qquad y \in Y.$$

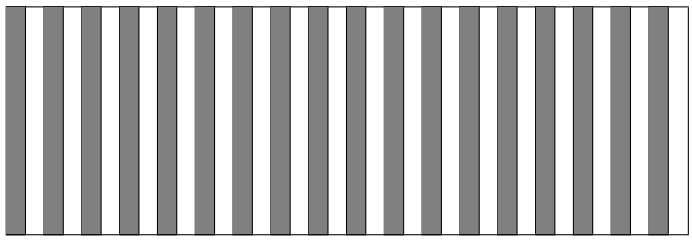
4. Pomocí funkcí χ^j lze k řešení u^* přidat tzv. korektor tak, že

$$U^\varepsilon(x) = u^*(x) - \varepsilon \left(\chi^1\Bigl(\frac{x}{\varepsilon}\Bigr)\,\frac{\partial u^*}{\partial x_1}(x) + \cdots + \chi^N\Bigl(\frac{x}{\varepsilon}\Bigr)\,\frac{\partial u^*}{\partial x_N}(x) \right)$$

vystihuje i lokální chování řešení $u^\varepsilon(x)$.

2.5. Příklad – vrstevnatý materiál

Ve speciálních případech lze homogenizované koeficienty odvodit intuitivně. Uvažujme dvojrozměrnou úlohu ustáleného vedení elektrického proudu, viz [8], v tenké desce složené ze dvou druhů A a B homogenního izotropního materiálu uspořádaného v úzkých svislých stejně širokých pruzích ve směru osy x_2, viz Obr. 2.



Obr. 2. Vrstevnatý materiál.

Neznámé elektrické napětí $u(x)$ potom bude splňovat rovnici

$$-\frac{\partial}{\partial x_1}\left(a\left(\frac{x_1}{\varepsilon}\right)\frac{\partial u}{\partial x_1} \right) - \frac{\partial}{\partial x_2}\left(a\left(\frac{x_1}{\varepsilon}\right)\frac{\partial u}{\partial x_2} \right) = f,$$

kde pravá strana f popisuje hustotu výkonu zdrojů proudu. Koeficient $a(x)$, který představuje elektrickou vodivost, nabývá střídavě dvou hodnot γ_A a γ_B. Jsou to převrácené hodnoty $\gamma_A = 1/\rho_A$ a $\gamma_B = 1/\rho_B$ elektrických odporů ρ_A, ρ_B. Pruhy vodivého materiálu lze považovat za elektrické „odpory“.

Pomocí známých výsledků z elementární fyziky o zapojení elektrických odporů lze odvodit „průměrné“ hodnoty odporu a vodivosti. Ve směru osy x_2 jsou tyto odpory „zapojeny“ paralelně, proto se sčítají jejich vodivosti. Celková vodivost ve směru osy x_2 je proto průměr obou vodivostí, tj. $b_2 = \frac{1}{2}(\gamma_A + \gamma_B)$. Ve směru osy x_1 jsou odpory „zapojeny“ sériově – odpory se sčítají, celkový odpor je průměr odporů, tj. $\frac{1}{2}(\rho_A + \rho_B)$. Převedením na vodivosti dostáváme koeficient

$b_1 = [\frac{1}{2}((\gamma_A)^{-1} + (\gamma_B)^{-1})]^{-1}$. Homogenizovaná rovnice potom je

$$-\frac{\partial}{\partial x_1}\left(b_1 \frac{\partial u}{\partial x_1}\right) - \frac{\partial}{\partial x_2}\left(b_2 \frac{\partial u}{\partial x_2}\right) \equiv -b_1 \frac{\partial^2 u}{\partial x_1^2} - b_2 \frac{\partial^2 u}{\partial x_2^2} = f,$$

kde homogenizované koeficienty jsou (obecně a v případě poměru složek 1 : 1)

$$b_1 = \left[\int_Y \frac{\mathrm{d}y}{a(y)}\,\mathrm{d}y\right]^{-1} = \left[\frac{1}{2}\left(\frac{1}{\gamma_A} + \frac{1}{\gamma_B}\right)\right]^{-1}, \quad b_2 = \int_Y a(y)\,\mathrm{d}y = \frac{1}{2}\left[\gamma_A + \gamma_B\right].$$

Pokud vodivosti složek jsou $\gamma_A = 1$ a $\gamma_B = 9$, potom b_1 je harmonický a b_2 aritmetický průměr vodivostí:

$$b_1 = \left[\frac{1}{2}\left(\frac{1}{1} + \frac{1}{9}\right)\right]^{-1} = \left[\frac{10}{2 \cdot 9}\right]^{-1} = 1.8 \qquad b_2 = \frac{1}{2}[1 + 9] = 5.$$

Homogenizací vznikl materiál anizotropní, i když obě jeho složky byly izotropní. Homogenizované koeficienty – efektivní vodivosti – tedy závisejí nejen na hodnotách složek a jejich poměru, ale i na geometrickém uspořádání složek. V případě jiného symetrického rozložení složek v periodě Y homogenizované koeficienty jsou různými „průměry“ koeficientu $a(y)$. Pokud rozložení složek v periodě není symetrické, matice homogenizovaných koeficientů b_{ij} může mít všechny členy nenulové.

3. Slabá konvergence

Posloupnost koeficientů $a_n(x) = a(nx)$ ve formulaci (2.3) problému homogenizace nekonverguje v obvyklém smyslu. Tato posloupnost však konverguje tzv. slabě.

3.1. Definice

Buď V reálný lineární prostor (zvaný také vektorový prostor), tj. množina V bodů u, na které jsou definovány operace sčítání $u+v$ a násobek reálným číslem αu, které splňují jisté axiomy, viz např. [7]. Budeme se zabývat normovaným prostorem, tj. lineárním prostorem V s normou. Připomeňme, že norma je funkcionál $\|\cdot\|$ na prostoru V s hodnotami v intervalu $\langle 0, \infty)$, který splňuje jistou podmínku homogenity $\|\alpha u\| = |\alpha| \cdot \|u\|$, trojúhelníkovou nerovnost $\|u + v\| \leq \|u\| + \|v\|$ a rozlišuje body: $\|u\| = 0 \Rightarrow u = 0$. Norma umožňuje zavést na V topologické pojmy jakým je okolí, otevřené a uzavřené množiny, spojitost, konvergentní a cauchyovské posloupnosti, úplnost, kompaktnost a další.

Na normovaném prostoru máme přirozenou konvergenci (zvanou také silná konvergence nebo konvergence v normě) definovanou pomocí normy

$$u_n \to u^* \qquad \Longleftrightarrow \qquad \|u_n - u^*\| \to 0\,.$$

Pro definici slabé konvergence potřebujeme pojem spojitého lineárního funkcionálu. Reálný funkcionál je zobrazení z prostoru V do reálných čísel $\mathbb{R}$. Funkcionál F na lineárním prostoru V nazveme lineární, jestliže zachovává operace součtu a skalárního násobku, tj. pro každé $u_1, u_2 \in V$ a každé $\alpha_1, \alpha_2 \in \mathbb{R}$ platí

$$F(\alpha_1 u_1 + \alpha_2 u_2) = \alpha_1 F(u_1) + \alpha_2 F(u_2).$$

Funkcionál F je spojitý, jestliže zachovává konvergenci

$$u_n \to u^* \qquad \Longrightarrow \qquad F(u_n) \to F(u^*)\,.$$

V případě lineárních funkcionálů je spojitost ekvivalentní omezenosti funkcionálu. Tyto funkcionály můžeme opět sčítat $F+G$ a dělat skalární násobky αF, proto množina všech spojitých lineárních funkcionálů na V tvoří opět lineární prostor, který značíme V^*, s „přirozenou“ normou

$$\|F\|_* = \sup_{\|u\|\le 1} |F(u)| \equiv \sup_{u\neq 0} \frac{|F(u)|}{\|u\|}\,, \tag{3.1}$$

ve které je V^* úplný normovaný prostor, tj. Banachův prostor. V definici slabé konvergence novou konvergenci testujeme pomocí těchto funkcionálů:

Definice 3.1. Posloupnost $\{u_n\}$ *konverguje slabě* k u^*, jestliže

$$F(u_n - u^*) \to 0 \qquad \text{pro všechny funkcionály} \quad F \in V^*\,.$$

Slabou konvergenci zapisujeme poloviční šipkou $u_n \rightharpoonup u^*$.

Definice (3.1) dává $F(x_n) - F(x^*) = F(x_n - x^*) \le \|F\|_* \cdot \|x_n - x^*\|$, odkud plyne, že každá (silně) konvergentní posloupnost konverguje i slabě. Obráceně to však neplatí, ne každá slabě konvergentní posloupnost konverguje i silně.

3.2. Prostor konečných vektorů $\mathbb{R}^N$

V prostoru $\mathbb{R}^N$ lze body ztotožnit s vektory $u \equiv (u_1, u_2, \dots, u_N)$ a jejich norma je $\|u\| = \left[\sum_{i=1}^N u_i^2\right]^{1/2}$. Podle definice posloupnost bodů $u_n \equiv (u_1^k, \dots, u_N^k)$ konverguje (silně) k $u^* \equiv (u_1^*, \dots, u_N^*)$, jestliže jejich vzdálenost jde k nule:

$$u_n \to u^* \qquad \Longleftrightarrow \qquad \|u_n - u^*\| \to 0\,.$$

Protože každý lineární funkcionál F na $\mathbb{R}^N$ lze napsat jako lineární kombinaci projekcí E_i na i-tou složku

$$E_i(u) \equiv E_i(u_1, \dots, u_i, \dots, u_N) = u_i\,,$$

k ověření slabé konvergence stačí vzít jen tyto projekce. Proto posloupnost vektorů $u_n = (u_1^n, \dots, u_N^n)$ *konverguje slabě* k u^*, pokud pro všechna i posloupnost $E_i(u_n) = u_i^n$ konverguje k $E_i(u^*) = u_i^*$, tj. každá složka konverguje:

$$u_n \rightharpoonup u^* \qquad \Longleftrightarrow \qquad u_1^n \to u_1^*,\ u_2^k \to u_2^*,\ \dots, u_N^n \to u_N^*.$$

A obráceně, pokud každá složka konverguje, posloupnost konverguje i v normě, protože těchto složek je konečně mnoho. Skutečně, jestliže pro $\varepsilon > 0$ jsou n_i indexy takové, že $|u_i^n - u_i^*| < \varepsilon$ pro každé $n > n_i$, potom pro $n > \max(n_1, \dots, n_N)$ platí

$$\|u_n - u^*\| = \left[\sum_{i=1}^n (u_i^n - u_i^*)^2\right]^{1/2} \le \left[\sum_{i=1}^n \varepsilon^2\right]^{1/2} = \varepsilon\sqrt{n}.$$

Přicházíme tak k důležitému výsledku:

Věta 3.2. *V normovaném prostoru konečné dimenze posloupnost konverguje slabě, právě když konverguje silně, tj. slabá a silná konvergence splývají.*

Další významnou vlastností je kompaktnost:

Věta 3.3. *Každá omezená posloupnost v normovaném prostoru konečné dimenze obsahuje konvergentní podposloupnost.*

Kompaktnost je velmi důležitá vlastnost, která se často využívá při důkazech existence řešení. Protože často nevíme, zda řešení nějaké rovnice existuje, konstruujeme posloupnost přibližných řešení. Pokud ukážeme, že tato posloupnost je omezená, díky kompaktnosti z ní vybereme konvergentní podposloupnost přibližných řešení, o jejíž limitě pak stačí ověřit, že je „přesným“ řešením dané rovnice.

3.3. Prostor nekonečných posloupností ℓ^2

Body prostoru ℓ^2 jsou nekonečné posloupnosti $u = (u_1, u_2, u_3, \dots)$, pro které platí

$$\|u\| = \left[\sum_{i=1}^{\infty} |u_i|^2\right]^{1/2} \leq \infty\,.$$

V prostoru ℓ^2 existují slabě konvergentní posloupnosti, které nekonvergují (silně):

Příklad 3.4. Omezená posloupnost $\{e_1, e_2, e_3, \dots\}$, tzv. „putující jednička“

$$e_1 = (1,0,0,0,\dots),\ e_2 = (0,1,0,0,\dots),\ e_3 = (0,0,1,0,\dots),\ \dots$$

není konvergentní, protože není cauchyovská: pro $i \neq j$ platí $\|e_i - e_j\| = \sqrt{2}$. Ukážeme, že je slabě konvergentní. Každý funkcionál F na prostoru ℓ^2 má tvar

$$F(u) = \sum_{i=1}^{\infty} f_i u_i = f_1 u_1 + f_2 u_2 + f_3 u_3 + \dots\,,$$

kde $f = (f_1, f_2, f_3, \dots)$ je také prvkem ℓ^2. Pro $n \to \infty$ platí

$$F(e_n) = f_1 \cdot 0 + \dots + f_{k-1} \cdot 0 + f_n \cdot 1 + f_{k+1} \cdot 0 + \dots = f_n\,.$$

Protože řada $\sum_{i=1}^{\infty} |f_i|^2$ je konvergentní, platí $f_n \to 0$. Proto posloupnost e_n konverguje slabě k nulové posloupnosti $0 = (0,0,0,\dots)$.

Příklad také ukazuje, že omezená posloupnost nemusí obsahovat konvergentní podposloupnost. Skutečně, posloupnost $\{e_n\}$ je omezená: $\|e_n\| = 1$, ale protože $\|e_i - e_j\| = \sqrt{2}$ pro $i \neq j$, žádná její podposloupnost nemůže být cauchyovská a proto ani konvergentní. V prostoru ℓ^2 však platí následující „užitečné“ tvrzení:

Věta 3.5. *Každá omezená posloupnost v ℓ^2 obsahuje slabě konvergentní podposloupnost, tj. uzavřená koule v ℓ^2 je kompaktní vzhledem ke slabé konvergenci.*

3.4. Prostor integrovatelných funkcí $L^2(I)$

Z hlediska aplikací je velmi důležitý tzv. Lebesgueův prostor, tj. prostor integrovatelných funkcí. Pro jednoduchost se budeme zabývat jenom případem funkcí na jednorozměrné oblasti Ω, tedy funkcemi na intervalu I. Prvky prostoru $L^2(I)$ jsou funkce $u(x)$ definované na I a mající konečnou normu $\|u\|$, tj. splňující

$$\|u\| = \left[\int_I |u(x)|^2 \,\mathrm{d}x\right]^{1/2} < \infty\,.$$

Prostor $L^2(I)$ je úplný normovaný prostor[2], tedy Banachův prostor.

Uveďme příklad posloupnosti funkcí konvergující slabě, ale nekonvergující silně.

[2]Pro úplnost dodejme, že integrál v definici je Lebesgueův integrál, který dovede integrovat i hodně nespojité funkce včetně Dirichletovy funkce. Funkce, které se liší jen v málo bodech (na množině míry nula), dávají stejný integrál, a uvedená norma je nerozliší. Proto tyto funkce

Příklad 3.6. Uvažujme Lebesgueův prostor $L^2(I)$ funkcí integrovatelných na intervalu $I = (0, 2\pi)$ a posloupnost $u_n(x) = \sin(nx)$. Tato posloupnost je omezená:

$$\|u_n\| = \left[\int_0^{2\pi} \sin^2(nx)\,\mathrm{d}x\right]^{1/2} = \left[\int_0^{2\pi} \frac{1}{2}(1 - \cos(2nx))\,\mathrm{d}x\right]^{1/2} = \sqrt{\pi}\,,$$

ale není konvergentní, protože není cauchyovská: snadno lze spočítat, že pro $n \neq m$

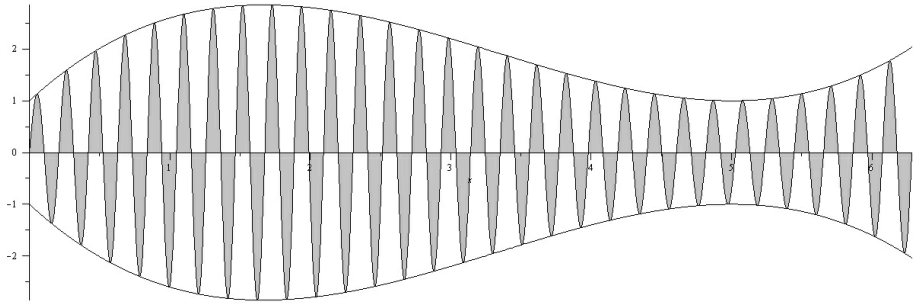
$$\|u_n - u_m\| = \left[\int_0^{2\pi} (\sin(nx) - \sin(mx))^2\,\mathrm{d}x\right]^{1/2} = \sqrt{2\pi}\,.$$

Ukážeme, že uvedená posloupnost konverguje slabě k nulové funkci. Lineární funkcionály na prostoru $L^2(I)$ mají tvar

$$F(u) = \int_I f(x)u(x)\,\mathrm{d}x\,,$$

kde f je funkce z $L^2(I)$. Potřebujeme dokázat, že

$$F(u_n) = \int_0^{2\pi} f(x)\sin(nx)\,\mathrm{d}x \to 0.$$



Obr. 3. Graf integrované funkce $f(x)\sin(nx)$ mezi grafy $-|f(x)|$ a $|f(x)|$.

Na Obr. 3 je graf součinu funkcí $f(x)\sin(nx)$. Je vidět, že se zmenšováním periody $2\pi/n$ oscilací přibývá, čímž se rozdíl mezi kladnými a zápornými částmi zmenšuje, proto integrál konverguje k nule.

Předchozí názorný obrázek však není důkaz. Naznačme ideu důkazu $u_n \rightharpoonup 0$. Využijeme přitom tvrzení:

Věta 3.7. *Nechť u_n je posloupnost funkcí na intervalu I omezená v normě prostoru $L^2(I)$. Jestliže pro každou hladkou funkci φ s nulami na koncích platí*

$$\int_I \varphi(x)(u_n(x) - u^*(x))\,\mathrm{d}x \to 0\,,$$

potom posloupnost u_n slabě konverguje k u^* v prostoru $L^2(I)$.

Posloupnost u_n je omezená v $L^2(I)$. Podle předchozí věty stačí dokázat konvergenci $\int_I f(x)u_n(x)\,\mathrm{d}x \to 0$ jen pro hladké funkce. Buď $f(x)$ diferencovatelná

ztotožníme. Prvky prostoru $L^2(I)$ jsou tedy třídy funkcí, které se liší jen v bodech z množiny míry nula, říkáme funkce, které se navzájem rovnají skoro všude. Je to také Hilbertův prostor.

funkce. Potom po integraci per-partes máme pro $n \to \infty$

$$\int_0^{2\pi} f(x)\sin(nx)\,\mathrm{d}x = \left[-f(x)\frac{\cos(nx)}{n}\right]_0^{2\pi} + \int_0^{2\pi} f'(x)\,\frac{\cos(nx)}{n}\,\mathrm{d}x \to 0\,,$$

díky členu $1/n$. Tím je slabá konvergence dokázána.

Předchozí příklad lze zobecnit na případ posloupnosti po částech konstantních koeficientů a^ε se zmenšující se periodou ε z homogenizace, tzv. „cimbuří“:

Příklad 3.8. Buď $\varphi(y)$ funkce z intervalu $\langle 0,1)$

$$\varphi(y) = \begin{cases} \xi_1 & \text{pro } y \in \langle 0,\lambda), \\ \xi_2 & \text{pro } y \in \langle \lambda,1) \end{cases}$$

periodicky rozšířená na celé $\mathbb{R}$. Pro $\varepsilon_n = \frac{1}{n} \to 0$ posloupnost funkcí

$$u_n(x) = \varphi(\tfrac{x}{\varepsilon_n}) \equiv \varphi(nx)$$

konverguje slabě v $L^2(I)$ ke konstantní funkci $u^*(x) = \lambda\xi_1 + (1-\lambda)\xi_2$.

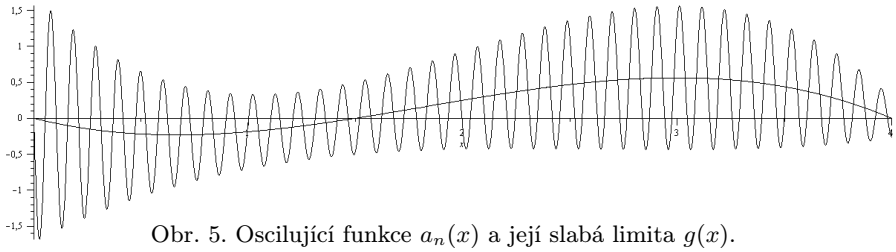
Obr. 4. Periodická funkce nabývající dvou hodnot, tzv. „cimbuří“.

Příklad zobecníme na posloupnost funkcí na oblasti Ω v $\mathbb{R}^N$ oscilujících s měnící se amplitudou $f(x)$ posunutých o $g(x)$:

Příklad 3.9. Buďte $f(x)$ a $g(x)$ funkce z $L^2(\Omega)$ a $\varphi(y)$ nenulová omezená integrovatelná Y-periodická funkce s nulovým průměrem, tj. $\int_Y \varphi(y)\,\mathrm{d}y = 0$. Potom pro $n \to \infty$ posloupnost funkcí oscilujících s amplitudou $f(x)$ posunutých o $g(x)$

$$a_n(x) = f(x)\,\varphi(nx) + g(x)$$

konverguje slabě v $L^2(\Omega)$ k funkci $g(x)$, nekonverguje však silně (kromě případu $f(x) = 0$). Oscilující část $f(x)\varphi(nx)$ nemá na slabou limitu vliv.



Obr. 5. Oscilující funkce $a_n(x)$ a její slabá limita $g(x)$.

Také v prostoru funkcí $L^2(\Omega)$ platí kompaktnost vzhledem ke slabé konvergenci:

Věta 3.10. *Každá posloupnost funkcí omezená v normě $L^2(\Omega)$ obsahuje slabě konvergentní podposloupnost.*

4. Problémy slabé konvergence

Ve srovnání se silnou konvergencí má slabá konvergence příjemnou vlastnost: umožňuje z omezené posloupnosti vybrat slabě konvergentní podposloupnost, má však některé nepříjemné vlastnosti: složení se spojitou funkcí nezachovává slabou konvergenci a v součinu dvou slabě konvergentních posloupností nelze přejít k součinu limit.

4.1. Složení se spojitou funkcí

Buď $\Phi : \mathbb{R} \to \mathbb{R}$ spojitá funkce splňující $|\Phi(\xi)| \leq k\,|\xi|$ $(k > 0)$. Snadno lze ověřit, že složením funkce $u \in L^2(I)$ s Φ dostaneme funkci $\Phi(u)$, která je opět v $L^2(I)$. Toto složení zachovává silnou konvergenci, tj. pro posloupnost $\{u_n\} \subset L^2(I)$ platí

$$u_n \to u^* \qquad \Longrightarrow \qquad \Phi(u_n) \to \Phi(u^*).$$

Slabá konvergence se však při složení nezachovává, implikace

$$u_n \rightharpoonup u^* \qquad \Longrightarrow \qquad \Phi(u_n) \rightharpoonup \Phi(u^*) \tag{4.1}$$

neplatí, jak je vidět z následujícího příkladu:

Příklad 4.1. Uvažujme funkci $\Phi(\xi) = \min(\xi^2, 2)$, která splňuje podmínku $|\Phi(\xi)| \leq k\,|\xi|$, a posloupnost $u_n(x) = \sin(nx)$ z Příkladu 3.6, která slabě konverguje v $L^2(I)$ k nulové funkci. Složená funkce

$$(\Phi(u_n))(x) = \sin^2(nx) = \tfrac{1}{2}(1 - \cos(2nx))$$

ale slabě konverguje k funkci $\frac{1}{2}$, která se nerovná $\Phi(\lim_{n\to\infty} u_n) = \Phi(0) = 0$.

Lze dokázat, že implikace (4.1) platí jen v případě lineární funkce $\Phi(\xi) = a\xi + b$.

4.2. Součin slabě konvergentních posloupností

Součin $w = u\,v$ funkcí $u, v \in L^2(I)$ je funkce z prostoru $L^1(I)$ funkcí, pro které $\int_I |w(x)|\,\mathrm{d}x < \infty$. Je to normovaný prostor s normou $\|w\|_1 = \int_I |w(x)|\,\mathrm{d}x$. Pro silně konvergující posloupnosti platí implikace:

$$u_n \to u^*\,, \quad v_n \to v^* \qquad \Longrightarrow \qquad u_n\,v_n \to u^* v^*,$$

kde na pravé straně je konvergence v normě $\|w\|_1$, a také konvergence integrálů

$$u_n \to u^*\,, \quad v_n \to v^* \qquad \Longrightarrow \qquad \int_I u_n v_n\,\mathrm{d}x \;\to\; \int_I u^* v^*\,\mathrm{d}x\,.$$

Obě implikace platí také v případě, kdy jedna z posloupností konverguje slabě:

$$u_n \to u^*\,, \quad v_n \rightharpoonup v^* \qquad \Longrightarrow \qquad u_n\,v_n \rightharpoonup u^* v^*.$$

V případě, kdy obě posloupnosti konvergují jenom slabě, implikace:

$$u_n \rightharpoonup u^*\,, \quad v_n \rightharpoonup v^* \quad \Longrightarrow \quad u_n\,v_n \rightharpoonup u^* v^* \tag{4.2}$$

neplatí, jak ukáže následující příklad:

Příklad 4.2. Uvažujme posloupnosti $u_n(x) = v_n(x) = \sin(nx)$ v prostoru $L^2((0, 2\pi))$. Obě posloupnosti slabě konvergují k nulové funkci, ale jejich součin

$$u_n(x)\,v_n(x) = \sin^2(nx) = \tfrac{1}{2}(1 - \cos(2nx))$$

slabě konverguje k funkci $\frac{1}{2}$, která se nerovná součinu slabých limit $0 \cdot 0$.

5. Závěr

Modelování kompozitních materiálů pomocí homogenizace vede na posloupnost úloh s koeficienty a^ε, které slabě konvergují. Ve slabé formulaci (2.4) úlohy (2.3) máme součin koeficientů a^ε a derivací $\frac{\partial u^\varepsilon}{\partial x_i}$ řešení u^ε. Protože posloupnost těchto parciálních derivací řešení u^ε je omezená v normě prostoru $L^2(\Omega)$, posloupnost parciálních derivací u^ε obsahuje slabě konvergentní podposloupnost $\frac{\partial u^{\varepsilon'}}{\partial x_i}$. Při $\varepsilon \to 0$ potřebujeme přejít k limitě v součinu dvou slabě konvergentních posloupností, což podle předchozích příkladů představuje problém. Řešení problému přechodu k limitě ve slabě konvergentní posloupnosti složené se spojitou funkcí a v součinu dvou slabě konvergentní posloupností bude obsahem pokračování tohoto článku.

Reference

[1] I. Babuška: *Homogenization approach in engineering*, v: Computing methods in applied sciences and engineering, Lecture notes in Econ. and Math. Systems 134, Springer, Berlin, 1976, 137-153.

[2] N. Bakhvalov, G. Panasenko: *Homogenization: averaging processes in periodic media*, Mathematical Problems in the Mechanics of Composite Materials, Springer, 1989.

[3] A. Bensoussan, J. L. Lions, G. Papanicolau: *Asymptotic analysis for periodic structures*, North Holland, 1978.

[4] D. Cioranescu, P. Donato: *Introduction to homogenization*, Oxford Univ. Press, USA, 2000.

[5] J. Franců: *Homogenizace – matematická metoda výpočtu materiálů s periodickou strukturou*, Stavebnícky časopis **31** (1983), 789–814.

[6] J. Franců: *Homogenizace*, sborník 6. seminář z parciálních diferenciálních rovnic – Manětín 1981, JČSMF a KMA VŠSE Plzeň 1982, 21–63.

[7] J. Franců: *Funkcionální analýza I*, skripta FSI VUT v Brně, Akad. nakl. CERM, Brno 2009.

[8] J. Franců: *Parciální diferenciální rovnice*, skripta FSI VUT v Brně, Akad. nakl. CERM, Brno 2011.

[9] J. Franců: *Moderní metody řešení diferenciálních rovnic*, skripta FSI VUT v Brně, Akad. nakl. CERM, Brno 2006.

Jan Franců, Ústav matematiky, Fakulta strojního inženýrství, Vysoké učení technické v Brně, Technická 2, 616 69 Brno, Česká republika,
e-mail: `francu@fme.vutbr.cz`

O KVATERNIONOVÝCH ALGEBRÁCH

LENKA MACÁLKOVÁ

Abstrakt. V tomto článku jsou prezentovány dvě ekvivalentní definice kvaternionových algeber a uvedeny některé příklady. Cílem je seznámit čtenáře se základními pojmy. Více podrobností lze nalézt v [1].

Úvod

V první části se seznámíme s pojmem kvaternionové algebry. Uvedeme dvě základní definice a ukážeme, že jsou ekvivalentní. Dále definujeme stopu a normu prvku. V další části poskytneme charakterizaci kvaternionových algeber. Inspirací pro tento článek byly příklady z knihy [1]. Jejich prostřednictvím se budeme čtenáři snažit problematiku alespoň částečně přiblížit.

1. Definice kvaternionových algeber

Nejprve uvedeme dvě definice kvaternionové algebry nad tělesem F. V první definici je třeba zvlášť ošetřit případ, kdy je těleso charakteristiky 2. Po zavedení dalších potřebných pojmů a uvedení potřebných tvrzení ukážeme, že jsou spolu ekvivalentní. V dalším textu budeme pak vycházet z obou definic a používat tu, která bude v danou chvíli pro naše potřeby vhodnější.

Definice 1.1 (První definice kvaternionové algebry)**.** Případ char $F \neq 2$: *Kvaternionovou algebrou* A nad tělesem F rozumíme vektorový prostor dimenze čtyři s bázovými vektory $1, i, j, k$, kde neutrálním prvkem vzhledem k operaci násobení na A je 1 a dále platí, že

$$i^2 = a, \quad j^2 = b, \quad ij = -ji = k,$$

kde $a, b \in F^*$.
Případ char $F = 2$: *Kvaternionovou algebrou* A nad F rozumíme, jako v předchozím případě, čtyřdimenzionální vektorový prostor s bází $1, i, j, k$, kde $1 \in F$ je neutrální prvek vzhledem k násobení. Narozdíl od prvního případu, zavedeme operaci násobení následujícím způsobem:

$$i^2 + i = a, \quad j^2 = b, \quad ij = j(1 + i) = k,$$

kde $a \in F, b \in F^*$.

2010 *MSC.* Primární 11R52.
Klíčová slova. kvaternionová algebra, norma, stopa.
Práce byla podporována projektem A-Math-Net – Síť pro transfer znalostí v aplikované matematice (CZ.1.07/2.4.00/17.0100).

Kvaternionovou algebru označujeme symbolem $\left(\frac{a,b}{F}\right)$ (tzv. *Hilbertův symbol*).

Poznámka 1.2. V dalším textu budeme uvažovat pouze tělesa, která nemají charakteristiku 2.

Než uvedeme druhou definici kvaternionové algebry, připomeneme pojmy, které se v definici budou vyskytovat.

Definice 1.3. Nechť R je okruh s jedničkou.

1. *Centrum* okruhu R je množina $Z(R) = \{a \in R | ra = ar,\ \forall r \in R\}$.
2. *R-algebra* A je okruh s jedničkou spolu s homomorfismem $f : R \longrightarrow A$, kde $f(1_R) = 1_A$ a $f(R) \subseteq Z(A)$.
3. A je *centrální R-algebra*, jestliže $f(R) = Z(A)$.
4. A je *jednoduchá R-algebra*, jestliže nemá žádné netriviální oboustranné ideály.

Poznámka 1.4. V dalším textu budeme v R-algebře A ztotožňovat okruh R s jeho obrazem $f(R)$.

Definice 1.5. Nechť F je těleso, $\overline{F}$ je jeho algebraický uzávěr a A je komutativní F-algebra. Pak A se nazývá *separabilní K-algebra*, jestliže $A \cong F[x]/(f(x))$, kde $f(x)$ je polynom s různými kořeny v $\overline{F}$.

Poznámka 1.6. Mějme těleso F a F-algebru A. Máme-li automorfismus ϕ na A, pro který platí, že ϕ zúženo na F je identita, budeme jej nazývat F-automorfismem.

Definice 1.7. *Konjugací* na F-algebře L, rozumíme zobrazení $m \mapsto \overline{m}$, $\forall m \in L$, které je netriviálním F-automorfismem nad L.

Definice 1.8 (Druhá definice kvaternionové algebry)**.** Nechť F je těleso. *Kvaternionová algebra* A s centrem F je centrální algebra A dimenze 4 nad F taková, že existuje separabilní algebra L dimenze 2 nad F a $\theta \in F^*$ tak, že $A = L + Lu$, kde $u \in A$ splňuje

$$u^2 = \theta, \quad um = \overline{m}u$$

pro všechna $m \in L$, kde $m \mapsto \overline{m}$ je konjugace na L. Tuto konjugaci rozšíříme na celou kvaternionovou algebru předpisem $\overline{u} = -u$.

Definice 1.9. Pomocí konjugace (viz definice 1.7) definujeme *normu* prvku $\mathrm{n}(x)$ a *stopu* $\mathrm{tr}(x)$ jako

$$\mathrm{n}(x) = x\overline{x}, \quad \mathrm{tr}(x) = x + \overline{x}.$$

Příklad 1.10. Prvky kvaternionové algebry A můžeme vnořit do okruhu matic nad $F(\sqrt{a})$. Je-li $\mathrm{char}\ F \neq 2$, stačí například uvážit vnoření:

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad i \mapsto \begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix} \quad j \mapsto \begin{pmatrix} 0 & b \\ 1 & 0 \end{pmatrix}.$$

Normu můžeme v tomto případě chápat jako determinant matice a stopu jako stopu matice.

Příklad 1.11. Nejznámějším příkladem kvaternionové algebry jsou tzv. *Hamiltonovy kvaterniony* $\mathcal{H}$. Jedná se o případ, kdy $F = \mathbb{R}$ a $i^2 = j^2 = k^2 = -1$, tj. $\mathcal{H} = \left(\frac{-1,-1}{\mathbb{R}}\right)$.

Mezi další příklady patří např. $A = \left(\frac{-2,3}{\mathbb{Q}}\right)$. Zde je $i^2 = a = -2$, $j^2 = b = 3$, $k^2 = -ab = 6$. Podívejme se nyní na násobení, normu a stopu v této algebře. Nechť $x = a_0 + a_1 i + a_2 j + a_3 k$, $y = b_0 + b_1 i + b_2 j + b_3 k$ jsou prvky algebry A. Pak

$$\begin{aligned} x \cdot y &= a_0b_0 - 2a_1b_1 + 3a_2b_2 + 6a_3b_3 + (a_0b_1 + a_1b_0 - 3a_2b_3 + 3a_3b_2)i + \\ &\quad +(a_0b_2 - 2a_1b_3 + a_2b_0 + +2a_3b_1)j + (a_0b_3 + a_1b_2 - a_2b_1 + b_3b_0)k. \end{aligned}$$

Tedy například

$$(1 + 2i - j + 3k)(1 - i + 2j + k) = 17 + 22i - 9j + 7k.$$

Pro výpočet stopy a normy budeme využívat konjugaci definovanou jako $\overline{x} = a_0 - a_1 i - a_2 j - a_3 k$. Potom

$$\begin{aligned} \mathrm{tr}(x) &= x + \overline{x} = 2a_0 \\ \mathrm{n}(x) &= x\overline{x} = a_0^2 + 2a_1^2 - 3a_2^2 - 6a_3^2. \end{aligned}$$

Například pro $x = 1+2i-j+3k$ je $\mathrm{tr}(x) = 2$, $\mathrm{n}(x) = -48$. Prvky této kvaternionové algebry můžeme reprezentovat maticemi následujícím způsobem:

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad i \mapsto \begin{pmatrix} \sqrt{-2} & 0 \\ 0 & -\sqrt{-2} \end{pmatrix} \quad j \mapsto \begin{pmatrix} 0 & 3 \\ 1 & 0 \end{pmatrix}.$$

Poznámka 1.12. Pro libovolné těleso F platí, že $M_2(F) \equiv \left(\frac{1,1}{F}\right)$. Pro izomorfismus stačí zvolit

$$1 \mapsto \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad i \mapsto \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad j \mapsto \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Příklad 1.13. Nechť A je kvaternionová algebra nad tělesem F. Mějme prvek $x = a_0 + a_1 i + a_2 j + a_3 k \in A \setminus F$. Nechť char $F \neq 2$. Pak snadno spočteme, že

$$x^2 = a_0^2 - aa_1^2 - ba_2^2 + aba_3^2 + 2a_0(a_1 i + a_2 j + a_3 k).$$

Tedy prvek $x \in A \setminus F$ je kořenem kvadratického polynomu $x^2 - x \cdot \mathrm{tr}(x) + \mathrm{n}(x)$.

Než ukážeme ekvivalenci definic 1.1 a 1.8, dokážeme některá pomocná tvrzení, která v důkazu využijeme.

Lemma 1.14. *Nechť F je těleso a $A = \left(\frac{a,b}{F}\right)$ je kvaternionová algebra podle definice 1.1. Pak platí*

1. $\left(\frac{a,b}{F}\right) \cong \left(\frac{ax^2,by^2}{F}\right)$ *pro libovolné* $x, y \in F^*$.
2. *A je centrální algebra.*
3. *A je jednoduchá algebra.*

Důkaz. 1. Mějme $A = \left(\frac{a,b}{F}\right)$ s bází $\{1,i,j,k\}$, $A' = \left(\frac{ax^2,by^2}{F}\right)$ s bází $\{1,i',j',k'\}$. Definujme zobrazení $\varphi : A' \longrightarrow A$ s předpisem $\varphi(1) = 1$, $\varphi(i') = xi$, $\varphi(j') = yj$ a $\varphi(k') = xyk$. Protože platí

$$\begin{aligned}(xi)^2 &= ax^2,\\ (yj)^2 &= by^2,\\ (xi)(yj) &= (xy)(ij) = -(xy)(ji) = -(yj)(xi),\end{aligned}$$

jedná se o izomorfismus F-algeber.

2. Nechť $\overline{F}$ je algebraický uzávěr F. Pak zřejmě platí $\left(\frac{a,b}{F}\right) \otimes_F \overline{F} \cong \left(\frac{a,b}{\overline{F}}\right)$. Protože $\overline{F}$ je algebraicky uzavřené, je každý jeho prvek čtvercem. Tedy podle první části tohoto tvrzení dostáváme, že $\left(\frac{a,b}{\overline{F}}\right) \cong \left(\frac{1,1}{\overline{F}}\right) \cong M_2(\overline{F})$. Centrem $M_2(\overline{F})$ je $\overline{F}$, a tudíž centrum $\left(\frac{a,b}{F}\right)$ je F.
3. Mějme nenulový ideál $I \subseteq A$ a nechť $\overline{F}$ opět značí algebraický uzávěr tělesa F. Protože $\left(\frac{a,b}{F}\right) \otimes_F \overline{F} \equiv \left(\frac{a,b}{\overline{F}}\right)$, pak $I \otimes_F \overline{F}$ je nenulový ideál v $M_2(\overline{F})$. Víme, že $M_2(\overline{F})$ je jednoduchá algebra. Navíc I je vektorový prostor nad F, který má dimenzi 4, a tedy $I = A$.

□

Nyní již ukážeme, že definice 1.1 a 1.8 jsou spolu ekvivalentní. Uvažme nejprve kvaternionovou algebru dle definice 1.8. Operaci násobení můžeme rozepsat následujícím způsobem:

$$(l_1 + l_2u)(l_3 + l_4u) = (l_1l_3 + l_2\overline{l_4}\theta) + (l_1l_4 + l_2\overline{l_3})u.$$

Protože L je separabilní algebra dimenze 2, je buď kvadratickým rozšířením tělesa F nebo je izomorfní součinu $F \times F$. V obou případech existují generátory L nad F, označme je $1, y$. Pak za bázi z definice 1.1 stačí zvolit $1, i = y, j = u$, tedy $b = \theta$.

Mějme nyní kvaternionovou algebru podle definice 1.1 s bází $1,i,j,k$. Tato algebra je podle lemmatu 1.14 centrální a jednoduchá. Rozšíření $F(i)$ je zřejmě separabilní, tedy $L = F(i)$. Položíme $u = j$ a ukážeme, že pak $A = L + Lu$ splňuje definici kvaternionové algebry 1.8. Stačí ověřit, že $um = \overline{m}u$ pro všechna $m \in L$.

Pokud je char $F \neq 2$, definujeme zobrazení $^-$ pro libovolné $x \in A$, $x = a_0 + a_1i + a_2j + a_3k$ jako

$$\overline{x} = a_0 - a_1i - a_2j - a_3k.$$

Prvky $m \in L$ jsou tedy ve tvaru $m = a_0 + a_1i$. Pak

$$j(a_0 + a_1i) = a_0j + a_1ji = a_0j - a_1ij = (a_0 - a_1i)j,$$

a tedy požadovaná rovnost platí. Pro char $F = 2$ definujeme

$$\overline{x} = (a_0 + a_1) + a_1i + a_2j + a_3k.$$

Potom dostáváme, že

$$j(a_0 + a_1i) = a_0j + a_1ji = a_0j + a_1j + a_1ij = ((a_0 + a_1) + a_1i)j.$$

2. Klasifikace kvaternionových algeber

Věta 2.1. *Každá kvaternionová algebra* $\left(\frac{a,b}{F}\right)$ *je buď okruh s dělením nebo je izomorfní okruhu matic* $M_2(F(\sqrt{a}))$.

Důkaz. Důkaz tvrzení pro náročnost není uveden. Lze jej nalézt například v [1] na straně 81. □

Poznámka 2.2. Okruhem s dělením rozumíme nekomutativní okruh, kde ke každému nenulovému prvku existuje prvek inverzní.

Uvažujme nyní kvaternionovou algebru dle definice 1.8.

Věta 2.3. *Nechť* $\mathrm{n}(L) = \{x \in F | x = \mathrm{n}(l)$ *pro* $l \in L\}$. *Pak* A *je izomorfní okruhu matic* $M_2(F(\sqrt{a}))$ *právě tehdy, když* L *není těleso nebo* $\theta \in n(L)$.

Důkaz. Než začneme s důkazem tvrzení, označme $i^2 = a$, $u^2 = \theta$. V prvním a druhém kroku ukážeme implikaci zprava doleva, ve třetím a čtvrtém kroku provedeme důkaz opačné implikace.

1. Nejprve ukážeme, že pokud L není těleso, pak A je izomorfní okruhu matic $M_2(F(\sqrt{a}))$. Nechť L není těleso. Pokud k prvku $a_0+a_1i \in L$ existuje prvek inverzní, je obecně ve tvaru $(a_0 + a_1i)^{-1} = \frac{a_0-a_1i}{a_0^2-aa_1^2}$. Protože L není těleso, existuje nenulový prvek a_0+a_1i, který inverzi nemá, tzn. $a_0^2-aa_1^2 = 0$. To ale znamená, že $a_0^2 = aa_1^2$. Uvážíme-li prvek $h \in A$, $h = (a_0+a_1i)+(a_0+a_1i)u$, pak

$$\begin{aligned} \mathrm{n}(h) &= (a_0 + a_1i + a_0u + a_1iu) \cdot (a_0 - a_1i - a_0u - a_1iu) = \\ &= (a_0^2 - aa_1^2) - (a_0^2 - aa_1^2)\theta = 0. \end{aligned}$$

   Prvek h má nulovou normu, a tedy je dělitelem nuly. Protože v A existují dělitelé nuly, je A je podle věty 2.1 izomorfní okruhu matic $M_2(F(\sqrt{a}))$.
2. Nechť nyní L je těleso a $\theta \in \mathrm{n}(L)$. Naším cílem je dokázat, že A je izomorfní okruhu matic. Protože $\theta \in \mathrm{n}(L)$, existuje $m = a_0 + a_1i \in L$ tak, že $\mathrm{n}(m) = a_0^2 - aa_1^2 = \theta$.

   Pro libovolný prvek $h \in A$ platí, že

$$h^2 = h_0^2 + h_1^2a + h_2^2\theta - h_3^2a\theta + 2h_0(h_1i + h_2u + h_3iu).$$

   Nyní ukážeme, že koeficienty h_0, h_1, h_2, h_3 lze zvolit tak, že $h^2 = 1, h \neq \pm 1$. Pokud zvolíme $h_0, h_1 = 0$, pak

$$h^2 = (h_2^2 - h_3^2a)\theta.$$

   Protože $\mathrm{n}(m) = \mathrm{n}(a_0+a_1i) = \theta$ a L je těleso, pak $\mathrm{n}(m^{-1}) = \mathrm{n}(\hat{a_0}+\hat{a_1}i) = \frac{1}{\theta}$. Položíme $h_2 = \hat{a_0}, h_3 = \hat{a_1}$, a tím jsme našli vhodný prvek $h \in A$. Pak

$$0 = h^2 - 1 = (h - 1)(h + 1),$$

   a tím jsme našli netriviální dělitele nuly a podle věty 2.1 je A izomorfní s $M_2(F(\sqrt{a}))$.

3. Ukažme nyní, že pokud A je izomorfní okruhu matic $M_2(F(\sqrt{a}))$ a současně L je tělesem, pak $\theta \in \mathrm{n}(L)$. Libovolný prvek $h \in A$ je tvaru $h = a_0 + a_1 i + (a_2 + a_3 i)u$, kde $a_k \in F$ pro $k = 0, \ldots, 3$. Pro normu platí

$$\mathrm{h}(x) = a_0^2 - a_1^2 a - a_2^2 \theta + a_3^2 a\theta = \mathrm{n}(m_1) - \theta \mathrm{n}(m_2),$$

kde $m_1, m_2 \in L$. Protože A je izomorfní okruhu matic, obsahuje dělitele nuly, a tedy existuje prvek $h \in A$ takový, že $\mathrm{n}(h) = 0$, tzn. $\mathrm{n}(m_1) - (\mathrm{n}(m_2))\theta = 0$. Protože L je těleso, pak $\theta = \frac{\mathrm{n}(m_1)}{\mathrm{n}(m_2)}$. Nyní ukážeme, že θ je normou konkrétního prvku z L. Nechť $m_1 = a_0 + a_1 i$, $m_2 = b_0 + b_1 i$, pak

$$\begin{aligned}
\theta &= \frac{a_0^2 - a_1^2 a}{b_0^2 - b_1^2 a} = \\
&= \frac{(a_0 + a_1 i)(a_0 - a_1 i)}{(b_0 + b_1 i)(b_0 - b_1 i)} = \\
&= \underbrace{\left(\frac{a_0 + a_1 i}{b_0 + b_1 i} \cdot \frac{b_0 - b_1 i}{b_0 - b_1 i}\right)}_{m} \underbrace{\left(\frac{a_0 - a_1 i}{b_0 - b_1 i} \cdot \frac{b_0 + b_1 i}{b_0 + b_1 i}\right)}_{\overline{m}} = \\
&= \mathrm{n}(m), \text{ kde } m \in L,
\end{aligned}$$

a tedy $\theta \in \mathrm{n}(L)$.

4. Nyní zbývá ukázat, je-li A izomorfní okruhu matic $M_2(F(\sqrt{a}))$ a současně $\theta \notin \mathrm{n}(L)$, pak L není těleso. Budeme postupovat sporem. Předpokládejme, že $A \equiv M_2(F(\sqrt{a}))$, $\theta \notin \mathrm{n}(L)$ a L je těleso. Pak se ale dostáváme do sporu, neboť podle 2. platí implikace

$$A \equiv M_2(F(\sqrt{a})) \text{ a současně } L \text{ je těleso } \Longrightarrow \theta \in \mathrm{n}(L).$$

□

Definice 2.4. Pokud je A kvaternionová algebra nad F izomorfní s $M_2(F(\sqrt{a}))$, pak říkáme, že A *se štěpí nad* F.

Poznámka 2.5. Z Weddeburnovy věty (viz [2] strana 556) víme, že každá konečná algebra s dělením je tělesem. Tedy každá kvaternionová algebra nad konečným tělesem F je izomorfní s $M_2(F)$.

Příklad 2.6. Uvažme nad tělesem $\mathbb{Q}$ dvě následující kvaternionové algebry

$$A = \left(\frac{-2, -3}{\mathbb{Q}}\right), \; A' = \left(\frac{2, 1}{\mathbb{Q}}\right).$$

Algebra A je algebrou s dělením, neboť pro normu prvků platí

$$\mathrm{n}(x) = a_0^2 + 2a_1^2 + 3a_2^2 + 6a_3^2,$$

a tedy jediným prvkem s nulovou normou je 0 a algebra A nemá dělitele nuly.

Algebra A' je izomorfní s $M_2(\mathbb{Q}(\sqrt{2}))$, neboť například $1 + i + j + k$ je dělitelem nuly, protože $\mathrm{n}(1 + i + j + k) = 0$.

Reference

[1] C. Maclachlan, A. W. Reid: *The Arithmetic of Hyperbolic 3-Manifolds*, Springer-Verlag, New York, 2003.
[2] D. S. Dummit, R. M. Foote: *Abstract Algebra*, 3rd ed., John Wiley & Sons, Hoboken, 2004.

Lenka Macálková, Ústav matematiky a statistiky, Přírodovědecká fakulta, Masarykova univerzita v Brně, Kotlářská 2, 611 37 Brno, Česká republika,
e-mail: macalkoval@gmail.com