Matrix semigroup homomorphisms from dimension two to three

Damjana Kokol-Bukovšek *

Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia

Received 28 August 1998; accepted 14 May 1999

Submitted by T.J. Laffey

Abstract

We characterise all non-degenerate homomorphisms from the multiplicative semigroup of all $2 \times 2$ matrices over an arbitrary field to the semigroup of $3 \times 3$ matrices over the same field. In the case of a field of real numbers every irreducible non-degenerate homomorphism is a conjugation of the symmetric square. © 1999 Elsevier Science Inc. All rights reserved.

Keywords: Matrix semigroup; Semigroup homomorphism; Multiplicative map; Reducibility; Field homomorphism

1. Introduction

Let $F$ be a field and let $\mathcal{M}_n(F)$ denote all $n \times n$ matrices with entries in $F$. In this paper we study matrix semigroup homomorphisms $\varphi : \mathcal{M}_2(F) \to \mathcal{M}_3(F)$, i.e., multiplicative maps. One way to obtain a semigroup homomorphism $\varphi : \mathcal{M}_n(F) \to \mathcal{M}_m(F)$ is to take a group homomorphism $\varphi' : GL_n(F) \to GL_m(F)$ and trivially extend it to all matrices taking $\varphi(A) = 0$ for every $A$ with $\det A = 0$. These trivial extensions are called degenerate and are known.

An example of non-degenerate semigroup homomorphism $\varphi : \mathcal{M}_n(F) \to \mathcal{M}_n(F)$ is the identity and an example of semigroup homomorphism $\varphi : \mathcal{M}_2(F) \to \mathcal{M}_3(F)$ is the symmetric square.
that is the mapping defined as follows. Let \( \mathcal{A} : \mathbb{F}^2 \to \mathbb{F}^2 \) be a linear transformation with a matrix \( A \). Then \( \text{Sym}^2 \mathcal{A} \) is a linear transformation from the symmetric tensor product \( \mathbb{F}^2 \otimes \mathbb{F}^2 \cong \mathbb{F}^3 \) to itself defined by

\[
(\text{Sym}^2 \mathcal{A})(x \vee y) = \mathcal{A}x \vee \mathcal{A}y
\]

and \( \text{Sym}^2 A \) is its matrix in the chosen basis. If \( \{e_1, e_2\} \) is the basis of \( \mathbb{F}^2 \), then

\[
\{e_1 \vee e_1, e_1 \vee e_2, e_2 \vee e_2\}
\]

is the corresponding basis of \( \mathbb{F}^2 \otimes \mathbb{F}^2 \) and

\[
\text{Sym}^2 \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} a^2 & ab & b^2 \\ 2ac & ad + bc & 2bd \\ c^2 & cd & d^2 \end{bmatrix}.
\]

We can obtain new semigroup homomorphisms \( \phi' : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_m(\mathbb{F}) \) from old ones \( \phi : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_m(\mathbb{F}) \), \( \varphi_1 : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_k(\mathbb{F}) \) and \( \varphi_2 : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_{m-k}(\mathbb{F}) \) by taking a direct sum, i.e.,

\[
\phi'(A) = \phi_1(A) \oplus \phi_2(A),
\]

by matrix conjugations, i.e.,

\[
\phi'(A) = S \phi(A) S^{-1}
\]

with \( S \) invertible and by using field homomorphism \( f : \mathbb{F} \to \mathbb{F} \) entrywise, i.e.,

\[
\phi'(A) = \left[ f(\phi(A)_{ij}) \right]_{i,j=1}^m.
\]

We will show that in case of semigroup homomorphisms from \( \mathcal{M}_2(\mathbb{F}) \) to \( \mathcal{M}_3(\mathbb{F}) \) and \( \text{char } \mathbb{F} \neq 2 \) these are all possibilities which can occur.

2. Preliminaries

We will first show that there is no loss of generality if we assume that a semigroup homomorphism \( \phi : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_m(\mathbb{F}) \) maps 0 to 0 and the identity to the identity.

Lemma 1. Let \( \mathbb{F} \) be a field and \( \phi : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_m(\mathbb{F}) \) a semigroup homomorphism. Then \( \phi \) has the form

\[
\phi(A) = S(\phi_0(A) \oplus E) S^{-1},
\]
where \( \varphi_0 : \mathcal{M}_n(\mathbb{F}) \to \mathcal{M}_k(\mathbb{F}) \) is a semigroup homomorphism with \( \varphi_0(0) = 0 \), \( \varphi_0(I) = I \), \( E \in \mathcal{M}_{m-k}(\mathbb{F}) \) is idempotent and \( S \in \mathcal{M}_m(\mathbb{F}) \) is an invertible matrix. Here either \( k \) or \( m-k \) may be 0, i.e., either \( \varphi_0(A) \) or \( E \) may be absent.

**Proof.** Since 0 and \( I \) are two commuting idempotents with \( 0I \equiv 0 \), \( u_0^\dagger \) and \( u_I^\dagger \) are also two commuting idempotents with \( u_0^\dagger u_I^\dagger \equiv u_0^\dagger \). So they have the form

\[
\begin{align*}
\varphi(0) &= S(0_k \oplus I_l \oplus 0_{m-l-k})S^{-1} \\
\varphi(I) &= S(I_k \oplus I_l \oplus 0_{m-l-k})S^{-1},
\end{align*}
\]

where \( 0_k, I_l \in \mathcal{M}_l(\mathbb{F}) \) and \( S \) is an invertible matrix. For any matrix \( A \in \mathcal{M}_n(\mathbb{F}) \) the matrix \( \varphi(A) \) commutes with \( \varphi(0) \) and \( \varphi(I) \), so it has the form

\[
\varphi(A) = S(A_1 \oplus A_2 \oplus A_3)S^{-1}.
\]

Since \( A0 = 0 \) and \( AI = A \) we have \( A_2I_l = I_l \) and \( A_30_{m-l-k} = A_3 \), so \( A_2 = I_l \) and \( A_3 = 0_{m-l-k} \). Writing \( \varphi_0(A) := A_1 \) we obtain the asserted form, since \( \varphi_0 \) is obviously a semigroup homomorphism. \( \square \)

In the proof of our main result we need the following proposition which is proved in [2].

**Proposition 1.** Let \( \mathbb{F} \) be a field and \( \varphi : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_2(\mathbb{F}) \) a semigroup homomorphism, which is non-degenerate and has the properties \( \varphi(0) = 0 \) and \( \varphi(I) = I \). Then \( \varphi \) has the form

\[
\varphi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = S\begin{bmatrix} f(a) & f(b) \\ f(c) & f(d) \end{bmatrix}S^{-1},
\]

where \( f : \mathbb{F} \to \mathbb{F} \) is a field homomorphism and \( S \in \mathcal{M}_2(\mathbb{F}) \) is an invertible matrix.

We will also need the following proposition, the proof of which is due to Radjavi.

**Proposition 2.** Let \( \mathbb{F} \) be a field and \( \varphi : \mathcal{M}_2(\mathbb{F}) \to \mathbb{F} \) a semigroup homomorphism. Then \( \varphi \) has the form

\[
\varphi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = h(ad - bc),
\]

where \( h : \mathbb{F} \to \mathbb{F} \) is a semigroup homomorphism.

**Proof.** If \( \varphi \) maps either everything to 0 or everything to 1, we take \( h \equiv 0 \) or \( h \equiv 1 \). If this is not the case, then \( \varphi(0) = 0 \) and \( \varphi(I) = 1 \). Matrix
is nilpotent so it is sent to 0. Every non-invertible matrix may be written as
\[ P \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} Q, \]
so it is sent to 0, too. Let us define \( h : \mathbb{F} \to \mathbb{F} \) as
\[ h(a) = \varphi \left( \begin{bmatrix} a & 0 \\ 0 & 1 \end{bmatrix} \right). \]
Then \( h \) is clearly a semigroup homomorphism. Every invertible matrix may be written as
\[ A = \begin{bmatrix} \det & A & 0 \\ 0 & 1 \end{bmatrix} A_1, \]
where \( A_1 \) has determinant 1. Thus we have to prove that every matrix of determinant 1 is sent to 1. Every \( 2 \times 2 \) matrix of determinant 1 is a product of two simple involutions, that is matrices similar to
\[ \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \]
if \( \text{char } \mathbb{F} \neq 2 \) and similar to
\[ \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \]
if \( \text{char } \mathbb{F} = 2 \) (see [3]). Homomorphism \( \varphi \) maps every simple involution to 1 or \(-1\). Since they are similar to each other they are all sent to 1 or all to \(-1\). In either case \( \varphi(A_1) = (\pm 1)^2 = 1. \) \( \square \)

3. Main result

The main result of this paper is the following:

**Theorem 1.** Let \( \mathbb{F} \) be a field and \( \varphi : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_3(\mathbb{F}) \) a semigroup homomorphism, which is non-degenerate and has the properties \( \varphi(0) = 0 \) and \( \varphi(I) = I. \) If \( \text{char } \mathbb{F} \neq 2 \) then \( \varphi \) has one of the following forms:

a. \[ \varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = S \begin{bmatrix} f(a) & f(b) & 0 \\ f(c) & f(d) & 0 \\ 0 & 0 & g(ad - bc) \end{bmatrix} S^{-1}, \]
where \( f : \mathbb{F} \to \mathbb{F} \) is a field homomorphism, \( g : \mathbb{F} \to \mathbb{F} \) is a semigroup homomorphism with \( g(0) = 0, g(1) = 1 \) and \( S \in \mathcal{M}_3(\mathbb{F}) \) is an invertible matrix,

\[
\phi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = S \begin{bmatrix} h(a^2) & h(ab) & h(b^2) \\ h(2ac) & h(ad + bc) & h(2bd) \\ h(c^2) & h(cd) & h(d^2) \end{bmatrix} S^{-1},
\]

where \( h : \mathbb{F} \to \mathbb{F} \) is a field homomorphism and \( S \in \mathcal{M}_3(\mathbb{F}) \) is an invertible matrix.

If \( \text{char} \mathbb{F} = 2 \) then \( \phi \) has one of the forms (a), (b) or

\[
\phi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = S \begin{bmatrix} h(a^2) & 0 & h(b^2) \\ h(ac) & h(ad + bc) & h(bd) \\ h(c^2) & 0 & h(d^2) \end{bmatrix} S^{-1},
\]

where \( h : \mathbb{F} \to \mathbb{F} \) is a field homomorphism and \( S \in \mathcal{M}_3(\mathbb{F}) \) is an invertible matrix.

**Remark 1.** If \( \text{char} \mathbb{F} = 2 \), cases (b) and (c) are essentially different: The image of \( \phi \) in case (b) has exactly one non-trivial invariant subspace in common, which has dimension 2. On the other hand, in case (c) the image of \( \phi \) has an invariant subspace of dimension 1 in common.

**Proof.** Let us denote by \( E_{ij} \) the matrix which has 1 in the \( i \)th row and the \( j \)th column, and 0 elsewhere. We will divide the proof in several steps.

**Step 1.** Without loss of generality we may assume that \( \phi(E_{12}) = E_{13} \) and \( \phi(E_{21}) = E_{31} \). Then \( \phi(E_{11}) = E_{11} \) and \( \phi(E_{22}) = E_{33} \).

**Proof.** Matrix \( E_{12} \) is nilpotent of order 2, so \( \phi(E_{12}) \) must be nilpotent of order at most 2. Let us suppose that \( \phi(E_{12}) = 0 \). If \( A \in \mathcal{M}_2(\mathbb{F}) \) is any non-invertible matrix, it has rank at most 1 and we can write it as \( A = PE_{12}Q \). So \( \phi(A) = \phi(P)\phi(E_{12})\phi(Q) = 0 \) and \( \phi \) is degenerate. Thus \( \phi(E_{12}) \) must be non-zero and we can write it as \( \phi(E_{12}) = xy^T \) where \( x, y \) are two column vectors in \( \mathbb{F}^3 \) and \( y^Tx = 0 \). Similarly we obtain \( \phi(E_{21}) = uv^T \) where \( v^Tu = 0 \). Since \( E_{12}E_{21}E_{12} = E_{12} \), we have

\[
xy^Tuv^Tx = xy^T,
\]

so \( y^Tu \cdot v^Tx = 1 \). With no loss of generality we may assume that \( y^Tu = v^Tx = 1 \).

Let us choose a vector \( z \in \mathbb{F}^3 \) orthogonal to \( v \) and \( y \), i.e., \( v^Tz = y^Tz = 0 \). Then \( \{x, z, u\} \) is a basis of \( \mathbb{F}^3 \). In this basis \( \phi(E_{12}) \) has the matrix \( E_{13} \) and \( \phi(E_{21}) \) has the matrix \( E_{31} \). So without loss of generality we may assume that \( \phi(E_{12}) = E_{13} \) and \( \phi(E_{21}) = E_{31} \). Then

\[
\phi(E_{11}) = \phi(E_{12}E_{21}) = E_{13}E_{31} = E_{11}
\]

and similarly \( \phi(E_{22}) = E_{33} \).
Step 2. \( \varphi(aI) \) has the form \( f(a)(E_{11} + E_{33}) + g(a)E_{22} \) where \( f, g : \mathbb{F} \to \mathbb{F} \) are semigroup homomorphisms with \( f(0) = g(0) = 0 \) and \( f(1) = g(1) = 1 \).

Proof. Matrix \( aI \) commutes with \( E_{12} \) and \( E_{21} \), so \( \varphi(aI) \) commutes with \( E_{13} \) and \( E_{31} \) and we obtain the asserted form.

Step 3. Homomorphism \( \varphi \) has the form

\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} f(a) & f(b) \\ * & * \\ f(c) & f(d) \end{bmatrix}
\]

Proof. If

\[
A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}
\]

is an arbitrary matrix, we have

\[
E_{11} \varphi(A)E_{11} = \varphi(E_{11}AE_{11}) = \varphi(aE_{11}) = \varphi(aI)E_{11} = f(a)E_{11},
\]

so the element in the first row and the first column of \( \varphi(A) \) must be \( f(a) \). Similarly we argue for the other corners.

Step 4. If \( A \) is upper-right (resp. upper-left, lower-right, lower-left) triangular, then \( \varphi(A) \) is upper-right (resp. upper-left, lower-right, lower-left) triangular. If \( A \) is diagonal, then \( \varphi(A) \) is diagonal.

Proof. Let

\[
A = \begin{bmatrix} a & b \\ 0 & d \end{bmatrix}
\]

Then

\[
\varphi(A)E_{11} = \varphi(AE_{11}) = \varphi(aE_{11}) = f(a)E_{11}
\]

and

\[
E_{33} \varphi(A) = \varphi(E_{22}A) = \varphi(dE_{22}) = f(d)E_{33}
\]

so the first column of \( \varphi(A) \) must be \( [f(a), 0, 0]^T \) and the last row must be \( [0, 0, f(d)] \). Thus \( \varphi(A) \) is upper-right triangular. Similarly we prove the other cases.

Step 5. If \( f(a) \neq g(a) \) for some \( a \in \mathbb{F} \), then

\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} f(a) & 0 & f(b) \\ 0 & h(ad - bc) & 0 \\ f(c) & 0 & f(d) \end{bmatrix},
\]
where $f : \mathbb{F} \to \mathbb{F}$ is a field homomorphism, $h : \mathbb{F} \to \mathbb{F}$ is a semigroup homomorphism, so we are in the case (a) of the theorem.

Proof. Matrix $aI$ commutes with every $A \in \mathcal{M}_2(\mathbb{F})$, so $\varphi(aI) = f(a)(E_{11} + E_{33}) + g(a)E_{22}$ commutes with $\varphi(A)$. Since $f(a) \neq g(a)$, $\varphi(A)$ has the form

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

Thus

$$
\varphi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} f(a) & 0 & f(b) \\
0 & s(a,b,c,d) & 0 \\
f(c) & 0 & f(d)\end{bmatrix}.
$$

So homomorphism $\varphi$ is a direct sum of two semigroup homomorphisms $\varphi_1 : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_2(\mathbb{F})$ and $\varphi_2 : \mathcal{M}_2(\mathbb{F}) \to \mathbb{F}$ where

$$
\varphi_1\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = \begin{bmatrix} f(a) & f(b) \\
f(c) & f(d)\end{bmatrix} \quad \text{and} \quad \varphi_2\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = s(a,b,c,d).
$$

Now, $f$ is a field homomorphism by Proposition 1 and $s(a,b,c,d)$ has the form $h(ad - bc)$ by Proposition 2.

From now on we will assume that $f(a) = g(a)$ for every $a \in \mathbb{F}$. So $\varphi(aI) = f(a)I$.

Step 6. If $\det A = 1$, then $\det \varphi(A) = 1$. Furthermore, $f(-1) = 1$ and

$$
\varphi(E_{12} - E_{21}) = E_{13} - E_{22} + E_{31}.
$$

Proof. Let $\varphi_1 : \mathcal{M}_2(\mathbb{F}) \to \mathbb{F}$ be the semigroup homomorphism $\varphi_1(A) = \det \varphi(A)$. By Proposition 2 it has the form $\varphi_1(A) = h(\det A)$. So, if $\det A = 1$, then $\det \varphi(A) = 1$. Now, $\det (-I) = 1$, so $\det (-I) = f(-1)^3 = 1$, thus $f(-1) = 1$. By step 4 $\varphi(E_{12} - E_{21})$ has the form $E_{13} + uE_{22} + E_{31}$. By the determinant condition we obtain $u = -1$.

From step 7 to step 14 we assume that $\text{char } \mathbb{F} \neq 2$.

Step 7. Without loss of generality we may assume

$$
\varphi(E_{11} + E_{12}) = E_{11} + E_{12} + E_{13}, \quad \varphi(E_{11} + E_{21}) = E_{11} + 2E_{21} + E_{31},
$$

$$
\varphi(E_{21} + E_{22}) = E_{31} + E_{32} + E_{33}, \quad \varphi(E_{12} + E_{22}) = E_{13} + 2E_{23} + E_{33}.
$$

Proof. Every matrix of rank one has the form $A = PE_{12}Q$ with $P,Q$ invertible. So its image has the form $\varphi(A) = \varphi(P)E_{13}\varphi(Q)$. Thus every matrix of rank 1 is sent to a matrix of rank 1. So the matrix $\varphi(E_{11} + E_{12})$ has rank 1. Since it is upper triangular, we have
\[ \phi(E_{11} + E_{12}) = E_{11} + xE_{12} + E_{13}, \]

Similarly
\[ \phi(E_{11} + E_{21}) = E_{11} + yE_{21} + E_{31}, \]
\[ \phi(E_{21} + E_{22}) = E_{31} + zE_{32} + E_{33}, \quad \phi(E_{12} + E_{22}) = E_{13} + tE_{23} + E_{33}. \]

Now,
\[
\begin{bmatrix}
  1 & x & 1 \\
  y & xy & y \\
  1 & x & 1
\end{bmatrix}
= \phi \left( \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) = \phi \left( \begin{bmatrix} 0 & 1 \\ 0 & 1 \end{bmatrix} \right)
= \begin{bmatrix}
  1 & z & 1 \\
  t & zt & t \\
  1 & z & 1
\end{bmatrix},
\]
so \( x = z \) and \( y = t \). Furthermore,
\[ \phi(0) = \phi \left( \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) = \begin{bmatrix} 2 - xy & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \]
so \( xy = 2 \). Since \( \text{char } \mathbb{F} \neq 2 \), both \( x \) and \( y \) are non-zero. If we take
\[ \phi'(A) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & 1 \end{bmatrix} \phi(A) \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1/x & 0 \\ 0 & 0 & 1 \end{bmatrix}, \]
we obtain
\[ \phi'(E_{11} + E_{12}) = E_{11} + E_{12} + E_{13}. \]

Homomorphism \( \phi' \) has all the properties we have proved for \( \phi \). So without loss of generality we may assume \( x = 1 \) and thus \( y = 2 \). (Actually we have multiplied the vector \( z \) from step 1 by a scalar, so we have chosen its length which was arbitrary in step 1.)

**Step 8.** \( \phi(E_{11} - E_{22}) = E_{11} - E_{22} + E_{33} \) and \( \phi(E_{12} + E_{21}) = E_{13} + E_{22} + E_{31}. \)

**Proof.** We have
\[ \phi(E_{11} - E_{22}) = E_{11} + vE_{22} + E_{33}, \]
so
\[ E_{11} + vE_{12} + E_{13} = (E_{11} + E_{12} + E_{13})(E_{11} + vE_{22} + E_{33}) \]
\[ = \phi((E_{11} + E_{12})(E_{11} - E_{22})) = \phi((E_{11} + E_{12})(E_{21} - E_{12})) \]
\[ = (E_{11} + E_{12} + E_{13})(E_{13} - E_{22} + E_{31}) = E_{11} - E_{12} + E_{13}. \]
Thus \( v = -1 \). Now,
\[
\varphi(E_{12} + E_{21}) = \varphi((E_{21} - E_{12})(E_{11} - E_{22})) = E_{13} + E_{22} + E_{31}.
\]

**Step 9.**
\[
\varphi\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \varphi\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix}.
\]

**Proof.** We have
\[
\varphi\left(\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}\right) = \begin{bmatrix} 1 & u & v & w \\ 0 & v & w & w \\ 0 & 0 & 1 \end{bmatrix}.
\]

Since
\[
\det \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} = 1, \quad v \text{ must be } 1.
\]

Furthermore,
\[
\begin{bmatrix} 1 & 1 & 1 \\ 2 & 2 & 2 \\ 1 & 1 & 1 \end{bmatrix} = \varphi\left(\begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}\right) = \varphi\left(\begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}\right)
\]
\[
= \begin{bmatrix} 1 & 1 & 1 \\ w & w & w \\ 1 & 1 & 1 \end{bmatrix},
\]

so \( w = 2 \). Similarly we prove \( u = 1 \) and the other equation.

**Step 10.** Mapping \( f : \mathbb{F} \to \mathbb{F} \) has the form \( f(a) = (h(a))^2 \), where \( h : \mathbb{F} \to \mathbb{F} \) is a semigroup homomorphism.

**Proof.** We have
\[
\varphi(aE_{11} + E_{22}) = f(a)E_{11} + h(a)E_{22} + E_{33},
\]
where \( h : \mathbb{F} \to \mathbb{F} \) is a semigroup homomorphism. Now,
\[
f(a)I = \varphi(aI) = \varphi((aE_{11} + E_{22})(E_{12} + E_{21})(aE_{11} + E_{22})(E_{12} + E_{21}))
\]
\[
= (f(a)E_{11} + h(a)E_{22} + E_{33})(E_{13} + E_{22} + E_{31})
\]
\[
\times (f(a)E_{11} + h(a)E_{22} + E_{33})(E_{13} + E_{22} + E_{31})
\]
\[
= f(a)E_{11} + h(a)^2E_{22} + f(a)E_{33}.
\]

So \( f(a) = h(a)^2 = (h(a))^2 \).

**Step 11.** \( \varphi(aE_{11} + bE_{22}) = h(a^2)E_{11} + h(ab)E_{22} + h(b^2)E_{33} \) and \( \varphi(aE_{12} + bE_{21}) = h(a^2)E_{13} + h(ab)E_{22} + h(b^2)E_{31} \).
Proof. If $b \neq 0$, we have

$$\varphi(aE_{11} + bE_{22}) = \varphi\left( bI\left( \frac{a}{b}E_{11} + E_{22} \right) \right) = f(b)f\left( \frac{a}{b} \right)E_{11} + f(b)h\left( \frac{a}{b} \right)E_{22} + f(b)E_{33} = h(a^2)E_{11} + h(ab)E_{22} + h(b^2)E_{33}$$

and

$$\varphi(aE_{12} + bE_{21}) = \varphi((aE_{11} + bE_{22})(E_{12} + E_{21})) = h(a^2)E_{13} + h(ab)E_{22} + h(b^2)E_{31}.$$ 

**Step 12.** Mapping $h : \mathbb{F} \to \mathbb{F}$ is a field homomorphism.

**Proof.** We have to prove that $h$ is additive.

$$\begin{bmatrix} h(a^2) & h(a(a + b)) & h((a + b)^2) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \varphi\left( \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & a + b \end{bmatrix} \right) = \varphi\left( \begin{bmatrix} 1 & 1 \\ 0 & 0 \\ 0 & b \end{bmatrix} \right) = \begin{bmatrix} h(a^2) & h(a^2) + h(ab) & h(a^2) + 2h(ab) + h(b^2) \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$ 

So $h(a(a + b)) = h(a^2) + h(ab)$. If $a \neq 0$, it follows $h(a + b) = h(a) + h(b)$.

**Step 13.**

$$\varphi\left( \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \right) = \begin{bmatrix} h(a^2) & h(ab) & h(b^2) \\ 0 & h(ad) & h(2bd) \\ 0 & 0 & h(d^2) \end{bmatrix}$$

and

$$\varphi\left( \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \right) = \begin{bmatrix} h(a^2) & 0 & 0 \\ h(2ac) & h(ad) & 0 \\ h(c^2) & h(cd) & h(d^2) \end{bmatrix}.$$ 

**Proof.** If $b \neq 0$, we have

$$\varphi\left( \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \right) = \varphi\left( \begin{bmatrix} 1 & 0 \\ 0 & d/b \end{bmatrix} \right) \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}.$$ 

$$\varphi\left( \begin{bmatrix} a & b \\ 0 & d \end{bmatrix} \right) = \varphi\left( \begin{bmatrix} 1 & 0 \\ 0 & d/b \end{bmatrix} \right) \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}.$$
\[
\begin{bmatrix}
1 & 0 & 0 \\
0 & d/b & 0 \\
0 & 0 & d^2/b^2
\end{bmatrix}
\begin{bmatrix}
1 & 1 & 1 \\
0 & 1 & 2 \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
h(a^2) & 0 & 0 \\
h(ab) & 0 & h(b^2)
\end{bmatrix}
= 
\begin{bmatrix}
h(a^2) & h(ab) & 0 \\
h(ab) & h(b^2) & h(d^2)
\end{bmatrix},
\]

Similarly we prove the other equation.

**Step 14.**

\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = 
\begin{bmatrix}
h(a^2) & h(ab) & h(b^2) \\
h(2ac) & h(ad + bc) & h(2bd) \\
h(c^2) & h(cd) & h(d^2)
\end{bmatrix},
\]

so we are in case (b) of the theorem.

**Proof.** If \(a \neq 0\), we have

\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \varphi \left( \begin{bmatrix} a & 0 \\ c & d - \frac{bc}{a} \end{bmatrix} \begin{bmatrix} 1 & b/a \\ 0 & 1 \end{bmatrix} \right)
= 
\begin{bmatrix}
h(a^2) & 0 & 0 \\
h(2ac) & h(ad - bc) & 0 \\
h(c^2) & h(cd - \frac{bc}{a}) & h((d - \frac{bc}{a})^2)
\end{bmatrix}
\begin{bmatrix}
1 & h(\frac{b}{a}) & h(\frac{b^2}{a}) \\
0 & 1 & h(\frac{b}{a}) \\
0 & 0 & 1
\end{bmatrix}
= 
\begin{bmatrix}
h(a^2) & h(ab) & h(b^2) \\
h(2ac) & h(ad + bc) & h(2bd) \\
h(c^2) & h(cd) & h(d^2)
\end{bmatrix}.
\]

If \(a = 0\) and \(d \neq 0\), then

\[
\varphi \left( \begin{bmatrix} 0 & b \\ c & d \end{bmatrix} \right) = \varphi \left( \begin{bmatrix} -\frac{bc}{d} & b \\ 0 & d \end{bmatrix} \begin{bmatrix} 1 & 0 \\ \frac{c}{d} & 1 \end{bmatrix} \right)
= 
\begin{bmatrix}
h(\frac{bc}{d^2}) & h(-\frac{bc}{d}) & h(b^2) \\
0 & h(-bc) & h(2bd) \\
0 & 0 & h(d^2)
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 \\
h(\frac{bc}{d^2}) & 1 & 0 \\
h(\frac{bc}{d}) & h(\frac{c}{d}) & 1
\end{bmatrix}
= 
\begin{bmatrix}
0 & 0 & h(b^2) \\
0 & h(bc) & h(2bd) \\
h(c^2) & h(cd) & h(d^2)
\end{bmatrix}.
\]

The case \(a = d = 0\) we have already proved in step 11.

**Step 15.** If \(\text{char } F = 2\), then either

\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = 
\begin{bmatrix}
h(a^2) & h(ab) & h(b^2) \\
0 & h(ad + bc) & 0 \\
h(c^2) & h(cd) & h(d^2)
\end{bmatrix}
\]
or
\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} h(a^2) & 0 & h(b^2) \\ h(ac) & h(ad + bc) & h(bd) \\ h(c^2) & 0 & h(d^2) \end{bmatrix}
\]
or
\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} h(a^2) & 0 & h(b^2) \\ 0 & h(ad - bc) & 0 \\ h(c^2) & 0 & h(d^2) \end{bmatrix},
\]
where \(h : \mathbb{F} \rightarrow \mathbb{F}\) is a field homomorphism, so we are in the cases (b), (c) or (a) of the theorem.

**Proof.** We do the same as in step 7 and obtain \(xy = 2 = 0\). If \(x \neq 0\), we may assume with no loss of generality that \(x = 1\) and then \(y = 0 = 2\). Then everything is the same as in steps 8–14 and we obtain the first possibility. If \(y \neq 0\), then we may assume with no loss of generality that \(y = 1\) and then \(x = 0 = 2\). In this case all the matrices in steps 8–14 are just transposed and we obtain the second possibility. If both \(x\) and \(y\) are 0, we obtain
\[
\varphi \left( \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \right) = \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}
\]
as in step 9 and
\[
\varphi \left( \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \right) = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}
\]
by the determinant condition. The semigroup \(\mathcal{M}_2(\mathbb{F})\) is generated by diagonal matrices and matrices
\[
\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}, \quad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}
\]
as we saw in steps 13, 14. So we obtain
\[
\varphi \left( \begin{bmatrix} a & b \\ c & d \end{bmatrix} \right) = \begin{bmatrix} f(a) & 0 & f(b) \\ 0 & h(ad - bc) & 0 \\ f(c) & 0 & f(d) \end{bmatrix}.
\]
Now \(\varphi(aI) = f(a)I\) gives that \(f(a) = h(a^2)\). Since \(f\) is additive by Proposition 1 and \(\text{char } \mathbb{F} = 2\), we have
\[(h(a + b))^2 = h((a + b)^2) = f(a + b) = f(a) + f(b) = h(a)^2 + h(b)^2 = (h(a) + h(b))^2,\]

so \(h\) is additive as well. \(\Box\)

4. Corollaries

Matrix semigroup homomorphism \(\phi\) is reducible if the image of \(\phi\) has a non-trivial invariant subspace in common, otherwise it is irreducible. We say that \(\phi\) is completely reducible if every invariant subspace of the image of \(\phi\) has an invariant complement.

**Corollary 1.** Let \(\mathbb{F}\) be a field with \(\text{char } \mathbb{F} \neq 2\). Every non-degenerate semigroup homomorphism \(\phi : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_3(\mathbb{F})\) is completely reducible.

**Corollary 2.** Let \(\phi : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_3(\mathbb{F})\) be an irreducible non-degenerate semigroup homomorphism. Then \(\text{char } \mathbb{F} \neq 2\) and

\[
\phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = S \begin{bmatrix} h(a^2) & h(ab) & h(b^2) \\ h(2ac) & h(ad + bc) & h(2bd) \\ h(c^2) & h(cd) & h(d^2) \end{bmatrix} S^{-1},
\]

where \(h : \mathbb{F} \to \mathbb{F}\) is a field homomorphism and \(S \in \mathcal{M}_3(\mathbb{F})\) is an invertible matrix.

If the field \(\mathbb{F}\) is the field of real numbers \(\mathbb{R}\), then the only non-zero field homomorphism of \(\mathbb{F}\) is the identity (see [1, p. 57]). This implies the following corollary.

**Corollary 3.** Let \(\phi : \mathcal{M}_2(\mathbb{R}) \to \mathcal{M}_3(\mathbb{R})\) be an irreducible non-degenerate semigroup homomorphism. Then

\[
\phi\left(\begin{bmatrix} a & b \\ c & d \end{bmatrix}\right) = S \begin{bmatrix} a^2 & ab & b^2 \\ 2ac & ad + bc & 2bd \\ c^2 & cd & d^2 \end{bmatrix} S^{-1},
\]

where \(S \in \mathcal{M}_3(\mathbb{R})\) is an invertible matrix.

If the field \(\mathbb{F}\) is the field of complex numbers \(\mathbb{C}\) we may be interested only in continuous semigroup homomorphism \(\phi : \mathcal{M}_2(\mathbb{F}) \to \mathcal{M}_3(\mathbb{F})\). Then semigroup or field homomorphisms \(f, g, h : \mathbb{F} \to \mathbb{F}\) in the Theorem 1 must be continuous. The only continuous field homomorphisms of \(\mathbb{C}\) are the identity and conjugation (see [1, p. 52]).
Acknowledgements

The author would like to thank Professor Matjaž Omladič for his helpful suggestions and Professor Heydar Radjavi for the proof of Proposition 2.

References