When are Two Elements of \( GL(2, \mathbb{Z}) \) Similar?

James T. Campbell

Department of Mathematical Sciences
Memphis State University
Memphis, Tennessee 38152

and

Elizabeth C. Trouy

Program in Applied Mathematics
Building 89
University of Arizona
Tuscon, Arizona 85721

Submitted by Lawrence Levy

ABSTRACT

We give an algorithm which determines conjugacy in \( GL(2, \mathbb{Z}) \), thus determining the topological conjugacy of the corresponding toral automorphisms.

0. INTRODUCTION

Whenever two elements \( g_1 \) and \( g_2 \) of a group \( G \) are related by a third element \( h \in G \) via the equation \( g_1 h = h g_2 \), \( g_1 \) and \( g_2 \) are said to be algebraically conjugate in \( G \), or just conjugate. If \( G \) is a group of matrices, then conjugacy is usually called similarity. It is well known that two elements of \( GL(n, \mathbb{R}) \) are similar precisely when they have the same Jordan form (given some ordering of the singular values). Here is an example of two members of \( GL(2, \mathbb{Z}) \) which have the same minimal and characteristic polynomials, yet are not conjugate over \( GL(2, \mathbb{Z}) \):

\[
A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 5 & 4 \\ 1 & 1 \end{bmatrix}.
\]
\[ x^2 - 6x + 1 \] is both the minimal and the characteristic polynomial for each of these, and they have the same Jordan and rational forms. If we suppose that for some \( C \) we have \( AC = CB \), then \( C \) must have the form

\[
C = \begin{bmatrix} a & 2b \\ b & 2(a - 2b) \end{bmatrix}.
\]

Assuming that \( a \) and \( b \) are integers, the determinant of \( C \) must be a multiple of 2, so that \( C \) cannot belong to \( \text{GL}(2, \mathbb{Z}) \).

Thus the Jordan form is not a complete similarity invariant in \( \text{GL}(2, \mathbb{Z}) \), and determining similarity in \( \text{GL} \) requires a deeper look into its structure. The purpose of this note is to give a self-contained description of an elementary algorithm which determines whether two elements of \( \text{GL}(2, \mathbb{Z}) \) are similar. (See [13] for a different account of this problem.) We do this by considering the problem in \( \text{SL}(2, \mathbb{Z}) \) (the subgroup of matrices in \( \text{GL} \) with determinant +1), and \( \text{PSL}(2, \mathbb{Z}) \). \( \text{PSL} \) is the quotient of \( \text{SL} \) by its center \( \{ \pm 1 \} \), i.e., each matrix is identified with its negative. Following the ideas laid out in [11], we obtain in Section II a presentation of \( \text{PSL} \) as a free product of cyclic groups. In such groups the solution to the conjugacy problem is fairly straightforward, as we describe in Section I. We are then able to extrapolate this information to obtain algorithms for determining conjugacy in \( \text{SL} \) and \( \text{GL} \) in Section III.

Besides being intrinsically interesting, the problem of similarity over \( \text{GL}(2, \mathbb{Z}) \) plays a central role in topological dynamics on the 2-torus. Section IV has more detailed comments on this.

I. ALGEBRAIC PRELIMINARIES

These are basic facts we need concerning free products of cyclic groups. They are standard fare.

**Definition.** A free product of cyclic groups is a group \( G \) with a presentation of the form \((x_1, \ldots, x_n; x_1^{r_1}, \ldots, x_n^{r_n})\), where the \( r_i \) are natural numbers.

In Section II we see how \( \text{PSL} \) may be presented as \((x, y; x^2, y^3)\).

If \( G \) has the presentation \((x_1, \ldots, x_n; x_1^{r_1}, \ldots, x_n^{r_n})\), a word \( W \) in \( G \) is reduced if \( W = x_i^{a_1} \cdots x_i^{a_n} \), where \( x_i \neq x_{i+1} \) and the \( a_k \) are positive integers modulo \( r_i \). A reduced word is one which has been written taking full advantage of the cyclic nature of the defining relators. There is an algorithm
for taking a given word and reducing it, thereby obtaining an equivalent word (i.e., a word which defines the same element of \( G \) as the original word). The algorithm is defined inductively on initial segments of increasing length in the word and agrees with our intuition, so we will not define it explicitly. Here is an example: in \((x, y; x^2, y^3)\), \( y^{-2}x^{-2}x^{-1}xyyx^{-1} \) is not reduced. We may delete \( x^{-2} \), and combine and delete \( x^{-1}x \), giving the word \( y^{-2}xyx^{-1} \). We combine \( y^{-2}y \) as \( y^{-1} \), then write the exponents in their nonnegative mod 2 and mod 3 equivalent forms to obtain \( y^2xyx \). This word is reduced and equivalent to the one we started with.

The following notion is a slightly restricted version of standard terminology. Given a word \( W = x_{n_1} \ldots x_{n_p} \) in the symbols \( x_{n_1}, \ldots, x_{n_p} \), the word \( W' \) is a cyclic permutation of \( W \) if there exists a \( j, 0 < j < p \), such that \( W' = x_{n_{j+1}} \ldots x_{n_p}x_{n_1} \ldots x_{n_j} \).

A word \( W = x_{n_1}^{a_1} \ldots x_{n_p}^{a_p} \) is cyclically reduced in a free product of cyclic groups if \( W \) is reduced and \( x_i \neq x_p \) for \( p \neq 1 \). The reader may provide examples of reduced words which are not cyclically reduced. However, if we take a reduced word \( W \), consider a cyclic permutation of it, reduce that word, then repeat this process on the newly reduced word, in a finite number of steps we will obtain a cyclically reduced word \( \sigma(W) \). This cyclically reduced word will not, in general, be unique. But every cyclically reduced word obtainable from \( W \) by this process will be a cyclic permutation of \( \sigma(W) \). Also \( \sigma(W) \) will not, in general, be equivalent to \( W \), but it will be conjugate to \( W \). For example, in \((x, y; x^2, y^3)\), \( W = yxy \) is reduced but not cyclically reduced (since \( yyx \) is not reduced). Either \( y^2x \) or \( xy^2 \) is a cyclic reduction for \( W \). These words are not equivalent to each other, but they are conjugate to each other and conjugate to \( W \). Anything conjugate to \( W \) can be cyclically reduced to one of them, and conversely. Here is the general theorem:

**Theorem 1.0.** Every element of a free product \( G \) of cyclic groups is defined by a unique word which is reduced in \( G \). Two words \( W_1 \) and \( W_2 \) are conjugate elements of \( G \) if and only if one of the cyclic reductions for \( W_1 \) is a cyclic permutation of one of the cyclic reductions for \( W_2 \).

Theorem 1.0 tells us how to solve the conjugacy problem given a presentation of a group as a free product of cyclic groups. Given two words, cyclically reduce them. If these cyclically reduced words are cyclic permutations of each other, then the two original words are conjugate; otherwise they are not.

For the rest of the paper \( G \) will denote the fixed group \((x, y; x^2, y^3)\). Theorem 1.0 allows us to determine all conjugacy classes in \( G \). We use the convention that the identity is the empty word. We agree to replace each
occurrence of \( y^2 \) in a reduced word by \( y^{-1} \). Thus every reduced word in \( G \) may be written as a word in the symbols \( x, y \), and \( y^{-1} \) with the property that if \( W = x_1^{n_1}x_2^{n_2}\ldots x_n^{n_n} \) is a word formed by these conventions, then (1) \( x_i \in \{x, y\} \) for each \( i \), (2) \( x_i \neq x_{i+1} \) for each \( i \), and (3) if \( x_i = x \) then \( \alpha_i = 1 \), and if \( x_i = y \) then \( \alpha_i = \pm 1 \).

The conjugacy classes will be determined when we find a representative from the cyclic permutation class of each cyclically reduced word of length \( n \), for \( n \in \mathbb{N} \). The cyclically reduced words of length 1 are \( x, y, y^{-1} \). For \( n = 2 \) we choose the representatives \( xy \) and \( xy^{-1} \). The other possible choices would be \( yx \) and \( y^{-1}x \), but these are cyclic permutations of the first two and hence their conjugacy classes are already represented. (For convenience we want \( x \) to appear first.)

There are no cyclically reduced words of length 3. Indeed, each reduced word of length 3 either begins and ends in \( x \), or in \( y \), or in \( y^{-1} \); or begins in \( y^2 \) and ends in \( y^{-1} \). None of these is cyclically reduced.

Proceeding inductively, we find that for \( n \geq 3 \) there are no cyclically reduced words of length \( n \) when \( n \) is odd, and if \( n = 2k \), each cyclically reduced word is a cyclic permutation of a word of the form

\[
(xy)^{r_1}(xy^{-1})^{r_2}(xy)^{r_3}\ldots(xy)^{r_{n-1}}(xy^{-1})^{r_n},
\]

where \( r_1, r_{j_k} \geq 0 \) and all other \( r_i \geq 1 \), and \( \sum_{i=1}^{k} r_i = k \).

Every word in \( G \) is conjugate to one of the words from the above list, and two words from \( G \) are conjugate precisely when their cyclic reductions are cyclic permutations of the same word from the above list.

II. SOME COMPLEX ARITHMETIC

Our goal in this section is to show that \( \text{PSL}(2, \mathbb{Z}) \) has the representation \( (x, y; x^2, y^3) \). The techniques are based upon several interesting exercises given in [11].

For each \( A \in \text{SL}(2, \mathbb{Z}) \),

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

let \( T_A \) denote the transformation of the complex plane \( \mathbb{C} \) given by \( T_A(z) = (az + b)/(cz + d) \). The set \( L \) of these transformations is a group under composition, and the surjective map \( U: \text{SL}(2, \mathbb{Z}) \to L \) given by \( U(A) = T_A \) is a
group homomorphism. It is easily checked that the kernel of $U$ is $\{ \pm I \}$; thus $L$ is isomorphic to $\text{PSL}(2, \mathbb{Z})$.

In $L$ we use complex arithmetic and elementary number theory to find some generators and relators. Let

$$W = -I, \quad X = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \quad \text{and} \quad Y = \begin{bmatrix} 0 & 1 \\ -1 & 1 \end{bmatrix};$$

then $T_x(z) = -1/z$, $T_y(z) = 1/(z - 1)$, and $(T_x T_y)^k(z) = z - k$ for all integers $k$. Thus for $A \in \text{SL}(2, \mathbb{Z})$,

$$T_x T_A(z) = \frac{-(cz + d)}{az + b} \quad \text{and} \quad (T_x T_y)^k T_A(z) = \frac{(a - kc)z + (b - kd)}{cz + d},$$

$k \in \mathbb{Z}$.

Now fix

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}.$$

We claim that by the following $T_A$ sufficiently often by $T_x$ and $(T_x T_y)^k$, we can obtain a transformation of the form $T(z) = (mz + j)/n$ with integers $m, n, j$ and $mn = 1$. If $c = 0$, then $T_A$ is already in this form, and if $a = 0$, a single composition on the left by $T_x$ puts $T_A$ in this form. We may assume that $a > |c|$ (otherwise compose on the left by $T_x$ and/or $T_w$), and apply the Euclidean algorithm. There is some $k_1 \in \mathbb{Z}$ such that $a = k_1 c + r_1$ where $r_1$ is an integral remainder, $0 < r_1 < |c|$. Thus

$$T_x (T_x T_y)^{k_1} T_A(z) = \frac{-(cz + d)}{r_1 z + j_1},$$

where $j_1 = b - k_1 d$. If $r_1 = 0$, we are done. Otherwise repeat this process, stopping when the remainder first becomes 0.

Since $mn = 1$, the resulting transformation $T(z)$ may be written in the form $T(z) = z + nj = (T_x T_y)^{-n}/(z)$.

For example, consider the transformation $T_A(z) = (5z + 2)/(2z + 1)$. Because $5 = (2)(2) + 1 = k_1 c + r_1$ and $j_1 = b - k_1 d = 0$, we compute $(T_x T_y)^2 T_A(z) = z/(2z + 1)$. Following by $T_x$ (and $T_w$, if you like), we get $T_x (T_x T_y)^2 T_A(z) = (2z + 1)/(-z)$. Since $2 = (-2)(-1) + 0 = k_1 c + r_1$ and
j_1 = b - k_1 d = 1, we compose with \((T_X T_Y)^{-2}\) and then by \(T_X\) to obtain finally \(T_X(T_X T_Y)^{-2}T_X(T_X T_Y)^{3}T_A(z) = z\).

Returning to the general case, applying this process to \(T_A\) shows that \(T_A\) may be written as a word in \(T_X\) and \(T_Y\). In fact,

\[
T_A(z) = T_X^{a_1}(T_X T_Y)^{-k_1}T_X^{a_2}(T_X T_Y)^{-k_2} \cdots (T_X T_Y)^{-k_p}T_X^{n_j}(T_X T_Y)^{-n_j}(z),
\]

where the \(a_i\)'s are either 0 or −1. For example, if

\[
A = \begin{bmatrix} 5 & 2 \\ 2 & 1 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 5 & 4 \\ 1 & 1 \end{bmatrix}
\]

are the two matrices from the introduction, we have \(T_A(z) = (T_X T_Y)^{-2}T_X^{-1}(T_X T_Y)^2T_X^{-1}(z)\) and \(T_B(z) = (T_X T_Y)^{-3}T_X^{-1}(T_X T_Y)^{-1}(z)\).

Thus \(L\) is generated by \(T_X\) and \(T_Y\), which satisfy \(T_X^2 = T_Y^3 = T_W\). Define a homomorphism from \(G = (x, y; x^2, y^3)\) to \(L\) by sending \(x\) to \(T_X\) and \(y\) to \(T_Y\) and extending appropriately. By the previous paragraph we see that this map is surjective. If a nonempty word in \(G\) gets sent to the identity in \(L\), then so does all of its conjugacy class. But direct computation (induced by induction) shows that none of the nonempty cyclically reduced words representing the conjugacy classes of \(G\), listed at the end of Section II, get sent to the identity. Thus our homomorphism is injective, so that \(G\) is isomorphic to \(L\) and hence to \(\text{PSL}(2, \mathbb{Z})\).

III. DETERMINANT, TRACE, AND CONJUGACY

We begin with \(\text{PSL}\). To determine the conjugacy of \(A\) and \(B\) in \(\text{PSL}(2, \mathbb{Z})\), first use the arithmetic algorithm from Section II to write \(T_A\) and \(T_B\) as words in \(T_X\) and \(T_Y\). Map these words into \(G\) via the isomorphism above, and using the conventions of Section I, freely reduce, then cyclically reduce, these words, and cyclically permute them until they are in one of the standard forms from the list at the end of Section I. If these are the same standard form, then \(A\) is conjugate to \(B\). Otherwise (by Theorem 1.0) it is not.

Applying this to the matrices \(A\) and \(B\) from the introduction, the cyclic reduction for \(A\) is \((x y)^2(x y^{-1})^2\) and the cyclic reduction for \(B\) is \((x y)(x y^{-1})^4\). They are not conjugate in \(\text{PSL}\).
Every element of $\text{SL}(2, \mathbb{Z})$ may be pulled back from PSL to a unique word $WV(X,Y)$ or $V(X,Y)$, where $V(X,Y)$ is a reduced word in $L$. Since

$$W^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

a presentation for $\text{SL}(2, \mathbb{Z})$ is $(X,Y; X^4, Y^3 = X^2)$. This is not a presentation as a free product of cyclic groups (there is an extra relation). But no matter; to solve the conjugacy problem in $\text{SL}(2, \mathbb{Z})$ we make two simple observations. First, if two matrices are conjugate in $\text{SL}$ then their images are conjugate in PSL. Second, there is only one conjugacy class in PSL with trace 0, and that is the class $(x)$.

Here is how to determine whether two elements $A$ and $B$ in $\text{SL}$ are conjugate in $\text{SL}$: Check the traces. If they're not equal, we're done; if they're both 0, see the next paragraph. If they're equal and nonzero, map them into PSL and check conjugacy there. By the observations just given, if their images are not conjugate in PSL, then $A$ and $B$ are not conjugate in $\text{SL}$; and if their images are conjugate, then $A$ is conjugate either to $B$ or to $-B$. But since the traces are equal and nonzero, we may conclude that $A$ must be conjugate to $B$.

The only cyclically reduced word in PSL with trace 0 is $x$. All matrices with trace 0 in $\text{SL}$ are conjugate to each other in $\text{SL}$ (easy and left to the reader).

If we again consider the matrices $A$ and $B$ from the introduction we note that this algorithm terminates when we map into PSL and discover they are not conjugate there, and hence not conjugate over $\text{SL}$.

The algorithm for $\text{SL}$ fails in $\text{GL}$ on two accounts. The first is that even within $\text{SL}$ it cannot see conjugation by matrices of determinant $-1$. For example, the algorithm tells us (correctly) that $XY$ and $XY^{-1}$ are not conjugate in $\text{SL}$; yet they are conjugate in $\text{GL}$ (by a matrix with determinant $-1$). The second is that it offers no information on how to determine conjugacy of two matrices each with determinant $-1$. The following two lemmas resolve these deficiencies. Let

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

**Lemma 3.1.** $A$ and $B$ in $\text{SL}$ are conjugate in $\text{GL}$ by a matrix of determinant $-1$ if and only if $A$ is conjugate to $ZBZ$ in $\text{SL}$.
Proof. If $\det C = -1$ then $C = ZC_1$ with $C_1 \in \text{SL}$. If $C$ satisfies $CA = BC$, then we have $C_1A = ZBZC_1$. The converse is immediate. \hfill \blacksquare

Lemma 3.2. Suppose $A$ and $B$ are $\text{GL}(2, \mathbb{Z})$. If $\det A = \det B$ and $\text{trace } A = \text{trace } B + 0$, then $A$ is conjugate to $B$ if and only if $A^2$ is conjugate to $B^2$.

Proof. $AC = CB$ implies $A(AC) = (CBC^{-1})(CB) = CB^2$. Conversely suppose $C^{-1}A^2C = B^2$. By assumption, the characteristic equations of $A$ and $B$ are the same [see (3.3)]. If

$$p(x) = \frac{1}{\text{trace } A} x + \frac{\det A}{\text{trace } A},$$

then $p(A^2) = A$ and $p(B^2) = B$. But $p(C^{-1}A^2C) = C^{-1}p(A^2)C = p(B^2)$, i.e., $AC = CB$. \hfill \blacksquare

The characteristic equation for $A \in \text{GL}(2, \mathbb{Z})$ is

(3.3) \hspace{1cm} x^2 - (\text{trace } A)x + \det A = 0.

This equation, and hence the trace and determinant, are conjugacy invariants.

Here is an algorithm for determining whether $B$ is conjugate to $A$ in $\text{GL}(2, \mathbb{Z})$: Check the traces. If they’re not equal, we’re done; if they’re both 0, see the next paragraph. If they’re equal and nonzero, check the determinants. If they’re not equal, we’re done; if they both equal 1, then apply the algorithm for $\text{SL}$. If they’re conjugate in $\text{SL}$, they’re conjugate in $\text{GL}$ and we’re done. If not, check the conjugacy of $A$ and $ZBZ$ in $\text{SL}$; if they’re conjugate in $\text{SL}$, then $A$ and $B$ are conjugate in $\text{GL}$, and if not, then $A$ and $B$ are not conjugate in $\text{GL}$; in either case we’re done. If the determinants both equal $-1$, apply this algorithm from the fourth sentence to $A^2$ and $B^2$; by Lemma 3.2 their conjugacy determines that of $A$ and $B$.

The trace 0 case requires considering the characteristic equation (3.3). If $\alpha_1$ and $\alpha_2$ are the roots of this equation, then $\alpha_1 + \alpha_2 = 0$ and $\alpha_1 \alpha_2 = \pm 1$. Hence $\alpha_1^2 = \pm 1$, and so $\alpha_1$ is either $\pm 1$ or $\pm i$. We observe that these three matrices include each of these cases:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \hspace{1cm} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \hspace{1cm} \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}.$$
We leave it to the reader to show that this exhausts all of the conjugacy classes in the trace 0 case.

Applying this algorithm to the matrices A and B of the introduction, we see that it terminates with a negative answer when A is not conjugate to $ZBZ$ in SL (details left to the reader).

IV. CLOSING REMARKS

The purely algebraic problem of similarity in $GL(n, \mathbb{Z})$ is equivalent to the a priori purely analytic problem of topological conjugacy of toral automorphisms. Given two maps $A$ and $B$ of a topological space $X$, they are topologically conjugate if there exists a homeomorphism $\phi$ of $X$ satisfying $A\phi = \phi B$. In other words, $A$ and $B$ are topologically conjugate if their actions on $X$ are the same up to a renaming of the points of $X$ by $\phi$. The torus is the topological group $\mathbb{R}^n / \mathbb{Z}^n$. A toral automorphism is a continuous map of the torus which is also a group automorphism. It is clear how each matrix in $GL(n, \mathbb{Z})$ induces a toral automorphism; moreover, each automorphism arises this way. (This can be seen by considering the induced action of the automorphism on the universal covering space $\mathbb{R}^n$.) In [2] and [4] it is shown that the topological conjugacy of the associated automorphisms is equivalent to the similarity of the inducing matrices, so that our algorithm may be applied for automorphisms of the 2-torus. For an excellent reference on the interesting geometry and dynamics in this situation we enthusiastically recommend [3]; see also [1], [6], and [7].

In his thesis [5], Ken Berg showed that a matrix in $GL(2, \mathbb{Z})$ may not be conjugate to its transpose, or its inverse. We leave these as exercises. The similarity problem is much more difficult in $GL(n, \mathbb{Z})$ for $n > 3$; for commentary see [9], and for the solution see [8] and [10].

The inspiration for this article came from the Master’s Thesis of Bruce Kitchens [9], which we read despite the title. We thank Cecil Rousseau for useful conversation.

REFERENCES


Received 5 July 1990; final manuscript accepted 3 August 1990