THE EQUATION $a^2b^2 = c^2$ IN FREE GROUPS

R. C. Lyndon

R. Vaught made the (unpublished) conjecture that *if* a, b *and* c *are elements of a free group and* $a^2b^2 = c^2$, *then* ab = ba. We establish this conjecture. Our method rests on an idea of J. Nielsen (see [3], also F. Levi [1], [3]), and it yields significant information about the solutions, in free groups, of the general 'quadratic equation,' in which each unknown appears, with exponent +1 or -1, at most twice. With simple modifications, into which we do not enter, our results carry over to free semigroups. The method does not apply fruitfully to equations beyond the quadratic, and no general method is known for deciding such questions as whether $a^3b^3 = c^3$ implies ab = ba.

1. PRELIMINARIES

Let w be an element of the group X with free generators $x_1, \cdots, x_n$. A *solution* of the ordered set $(w; x_1, \cdots, x_n)$ is an ordered set $(\phi; y_1, \cdots, y_m)$, where $y_1, \cdots, y_m$ are free generators of a group Y and ϕ is a homomorphism of X into Y such that $\phi w = 1$; ordinarily there is no ambiguity in saying more simply that ϕ is a solution of w. Informally, if we write

$$w = w(x_1, \cdots, x_n), \qquad \phi x_\nu = u_\nu(y_1, \cdots, y_m),$$

this expresses the fact that

$$w(u_1(y_1, \cdots, y_m), \cdots, u_n(y_1, \cdots, y_m)) = 1.$$

If ϕ is any homomorphism of a free group X of rank n into a second free group Y, it follows from a theorem of Nielsen [2] that the image ϕX is a free group of rank r no greater than n; the *rank* of ϕ is r, and the *nullity* is n - r. If ϕ: X into Y is a solution of w, and θ is a homomorphism of Y into a further free group Y', then $\theta\phi$: X into Y' is also a solution of w, of rank no greater than that of ϕ; we call $\theta\phi$ a *specialization* of ϕ. If T is any endomorphism of X, and ϕ a solution of Tw, then $\phi Tw = 1$ and ϕT is a solution of w. A set of solutions is *complete* if every solution is a specialization of some member of the set.

(*Remark.* The maximal rank of a solution of w may be generalized by defining the 'inner rank' of an arbitrary group to be the upper bound of the ranks of free homomorphic images.)

Each element w in X has a unique representation by a *reduced word*, that is, a formal product

$$x_{\nu_1}^{\varepsilon_1} \cdots x_{\nu_t}^{\varepsilon_t} \qquad (t \geq 0;\ \nu_i = 1, 2, \cdots, n;\ \varepsilon_i = \pm 1)$$

Received February 13, 1958.

This work was done in part while the author was associated with National Science Foundation projects, Grants G-2371 and G-1350.

that is reduced in the sense that, for no i, is $\nu_i = \nu_{i+1}$ and $\varepsilon_i = -\varepsilon_{i+1}$. Suppose, for some $h = 1, 2, \cdots, t-1$, that $\nu_h \neq \nu_{h+1}$, and define an automorphism R of X by

$$(1) \qquad Rx_{\nu_h} = \left(x_{\nu_h}^{\varepsilon_h} x_{\nu_{h+1}}^{-\varepsilon_{h+1}}\right)^{\varepsilon_h} \quad \text{or} \quad (1') \qquad Rx_{\nu_{h+1}} = \left(x_{\nu_h}^{-\varepsilon_h} x_{\nu_{h+1}}^{\varepsilon_{h+1}}\right)^{\varepsilon_{h+1}},$$

with $Rx_\nu = x_\nu$ for all other x_ν. Then R will be called an *elementary regular transformation belonging to* w. For each $h = 1, 2, \cdots, t$, the endomorphism S defined by

$$(2) \qquad Sx_{\nu_h} = 1, \quad \text{with } Sx_\nu = x_\nu \text{ for all other } x_\nu,$$

is an *elementary singular transformation belonging to* w. An endomorphism T of X *belongs to* w if it is a product $T = T_p T_{-p-1} \cdots T_1$ $(p \geq 0)$, where each T_i is an elementary transformation, regular or singular, belonging to $T_{i-1} T_{i-2} \cdots T_1 w$. Evidently the nullity of T is the number of singular T_i.

2. THE NIELSEN ARGUMENT

The *length* of a word

$$a = y_{\mu_1}^{\alpha_1} \cdots y_{\mu_s}^{\alpha_s}$$

in the generators $y_1, \cdots, y_m$ is $L(a) = s$. The length of an element w is that of the reduced word a representing the element: $L(w) = L(a)$.

With each word a we associate a sequence of integers

$$S(a) = (s, \mu_1, \alpha_1, \mu_s, -\alpha_s, \mu_2, \alpha_2, \mu_{s-1}, -\alpha_{s-1}, \cdots, \mu_1, -\alpha_1),$$

and define $H(a) = 1, 2, 3, \cdots$ to be the lexicographic rank of S(a) in the set of S(a') for all words a'. For an element w of Y (or a word representing w) we define $K(w) = \min(H(a), H(a^{-1}))$, where a is the reduced word representing w. Evidently K ranks the elements of Y in a manner compatible with their lengths, and it orders the finite set of elements of a given length in such a way that two elements have the same rank if and only if they are either equal or inverse to each other.

LEMMA 1. *Suppose that the products*

$$a = y_{\mu_1}^{\alpha_1} \cdots y_{\mu_s}^{\alpha_s}, \qquad b = y_{\nu_1}^{\beta_1} \cdots y_{\nu_s}^{\beta_s}, \qquad c = y_{\rho_1}^{\gamma_1} \cdots y_{\rho_t}^{\gamma_t},$$

ac, *and* bc *are reduced, and that* $s \leq t$. *If the sequence* $(\mu_1, \alpha_1, \mu_2, \alpha_2, \cdots, \mu_s, \alpha_s)$ *precedes lexicographically the sequence* $(\nu_1, \beta_1, \nu_2, \beta_2, \cdots, \nu_s, \beta_s)$, *then* $K(ac) < K(bc)$.

Proof. Since $s \leq t$, S(ac) has an initial segment

$$(s + t, \mu_1, \alpha_1, \rho_t, -\gamma_t, \cdots, \rho_{t-s+1}, -\gamma_{t-s+1}, \mu_s, \alpha_s).$$

The hypothesis on the sequences of indices from a and b implies that this segment precedes the corresponding segment from S(bc), whence $H(ac) < H(bc)$. The same considerations show that $H(c^{-1}a^{-1}) < H(c^{-1}b^{-1})$, whence $K(ac) < K(bc)$.

LEMMA 2. *Let* $v_1, \cdots, v_t$ *be elements of* Y $(t > 0)$, *no* $v_i = 1$, *and* $v_1 \cdots v_t = 1$. *Then, for some* $i = 1, 2, \cdots, t - 1$, *either* $K(v_i v_{i+1}) < K(v_i)$ *or* $K(v_i v_{i+1}) < K(v_{i+1})$.

Proof. Assume that, by way of contradiction,

$$K(v_i v_{i+1}) \geq K(v_i), K(v_{i+1}) \quad \text{for all } i = 1, 2, \cdots, t - 1. \tag{3}$$

We may suppose the v_i represented by reduced words; then the reduced word representing $v_i v_{i+1}$ will be obtained by cancelling some number $k \geq 0$ of letters y_ν^ε of the word representing v_i against an equal number from that representing v_{i+1}. Thus $L(v_i v_{i+1}) = L(v_i) + L(v_{i+1}) - 2k$, and, by (3),

$$k \leq \frac{1}{2}L(v_i), \frac{1}{2}L(v_{i+1}). \tag{4}$$

If a_i is the initial segment of the word representing v_i that cancels in forming the product $v_{i-1}v_i$, and a_{i+1}^{-1} the final segment that cancels in $v_i v_{i+1}$, these segments can not overlap, and we can represent v_i by a reduced product $v_i = a_i b_i a_{i+1}^{-1}$, where conceivably $b_i = 1$. In the extreme cases we write $v_1 = b_1 a_2^{-1}$, $v_t = a_t b_t$.

We show that no $b_i = 1$. That $b_1, b_t \neq 1$ follows from (4). Suppose that $b_i = 1$ for some $i = 2, 3, \cdots, t - 1$. Necessarily a_i and a_{i+1}^{-1} are the two equal halves of v_i, and, simplifying notation, we may write $v_{i-1} = da^{-1}$, $v_i = ab^{-1}$, $v_{i+1} = bc$, reduced words, with $L(a) = L(b)$. If $L(c) < L(b)$, then $L(v_i v_{i+1}) = L(ac) < L(ab) = L(v_i)$, contrary to (3); using a similar argument, we have $L(a) = L(b) \leq L(c), L(d)$. Since $v_i \neq 1$, $a \neq b$. If, as in Lemma 1, the sequence for a precedes that for b, it follows from the lemma that $K(v_i v_{i+1}) = K(ac) < K(bc) = K(v_{i+1})$, contrary to (3), while the reverse order gives $K(v_{i-1} v_i) = K(bd^{-1}) < K(ad^{-1}) = K(v_{i-1})$, again contrary to (3). Thus $b_i = 1$ contradicts the hypotheses.

But $v_1 \cdots v_t = b_1 a_2^{-1} a_2 b_2 a_3^{-1} \cdots a_t b_t = b_1 b_2 \cdots b_t$. Each $b_i \neq 1$, and each $a_i b_i b_{i+1} a_{i+2}^{-1}$ is reduced by hypothesis, whence each b_i has a last letter and b_{i+1} a first letter that do not cancel against one another. It follows that $b_1 b_2 \cdots b_t$ is reduced, and not 1 since $t > 0$, which contradicts the hypothesis that $v_1 \cdots v_t = 1$.

LEMMA 3. *If* $w \neq 1$ *and* ϕ *is a solution of* w, *then either* $\phi x_\nu = 1$ *for some* $\nu = 1, 2, \cdots, n$, *or else there is an elementary regular transformation* R *belonging to* w *such that*

$$\sum_{\nu=1}^{n} K(\phi R^{-1} x_\nu) < \sum_{1}^{n} K(\phi x_\nu).$$

Proof. Write $\phi x_\nu = u_\nu$. Then

$$\phi w = u_{\nu_1}^{\varepsilon_1} \cdots u_{\nu_t}^{\varepsilon_t} = 1 \qquad (t > 0).$$

If no $u_\nu = 1$, by Lemma 2 we may suppose, by symmetry, that

$$K\left(u_{\nu_h}^{\varepsilon_h} u_{\nu_{h+1}}^{\varepsilon_{h+1}}\right) < K\left(u_{\nu_h}^{\varepsilon_h}\right),$$

for some $h = 1, 2, \cdots, t$. Then, for R defined by (1), we have

$$K\left(\phi R^{-1} x_{\nu_h}\right) = K\left(\phi\left(x_{\nu_h}^{\varepsilon_h} x_{\nu_{h+1}}^{\varepsilon_{h+1}}\right)\right) = K\left(u_{\nu_h}^{\varepsilon_h} u_{\nu_{h+1}}^{\varepsilon_{h+1}}\right) < K\left(u_{\nu_h}^{\varepsilon_h}\right) = K(\phi x_{\nu_h}),$$

while $K(\phi R^{-1} x_\nu) = K(\phi x_\nu)$ for all other x_ν.

LEMMA 4. *If* $w \neq 1$ *and* ϕ *is a solution of* w, *then there is a regular transformation* T *belonging to* w *such that* $\phi T^{-1} x_\nu = 1$ *for some* $\nu = 1, 2, \cdots, n$.

Proof. If some $\phi x_\nu = 1$, we may take T to be the identity. We proceed by induction on $k = \Sigma K(\phi x_\nu)$. For small k, $k < 2n$, some $K(\phi x_\nu) = 1$ and $\phi x_\nu = 1$. Inductively, suppose the conclusion established for all $k' < k$, and that no $\phi x_\nu = 1$. By Lemma 3, there is an R belonging to w such that $k' = \Sigma K(\phi R^{-1} x_\nu) < \Sigma K(\phi x_\nu) = k$. Since $\phi' = \phi R^{-1}$ is a solution of $w' = Rw$, the induction hypothesis gives a T belonging to w' such that some $\phi' T^{-1} x_\nu = 1$. But then TR is a regular transformation belonging to w such that $\phi(TR)^{-1} x_\nu = \phi R^{-1} T^{-1} x_\nu = \phi' T^{-1} x_\nu = 1$.

LEMMA 5. *If* ϕ *is a solution of* w, *there exists a transformation* T *belonging to* w *such that* $Tw = 1$ *and* ϕ *is a specialization of* T.

Proof. If $w = 1$, taking T to be the identity gives $\phi = \phi T$. We proceed by induction on n, the rank of X. If $n = 0$, necessarily $w = 1$. Suppose that the conclusion is established for all $n' < n$, and that $w \neq 1$. If some $\phi x_\mu = 1$, where x_μ does not appear in w, then w lies in X' generated by the remaining x_ν, and the restriction ϕ' of ϕ to X' is a solution of w. The induction hypothesis gives an endomorphism T' of X' belonging to w such that $T'w = 1$ and $\phi' = \theta' T'$ for some θ' from X' into Y. Extending T' to T by $Tx_\mu = x_\mu$, and θ' to θ by $\theta x_\mu = 1$ gives the desired conclusion. If some $\phi x_{\nu_h} = 1$, then S, defined by (2), belongs to w, and $\phi = \phi S$ is a solution of Sw. Since Sw does not contain x_{ν_h}, the previous case gives $T(Sw) = 1$ and $\phi = \theta T$, whence $(TS)w = 1$ and $\phi = \theta TS$. Finally, if no $\phi x_\nu = 1$, Lemma 4 gives $\phi R^{-1} x_\nu = 1$, and, since ϕR^{-1} is a solution of Rw, the previous cases gives $\phi R^{-1} = \theta T$ and $T(Rw) = 1$, whence $\phi = \theta TR$ and $(TR)w = 1$.

Lemma 5 may be reformulated as follows.

PROPOSITION 6. *The set of transformations* T *belonging to* w *such that* $Tw = 1$ *constitutes a complete set of solutions of* w.

If w has solutions of nullity 0, then there is a transformation T, of nullity 0, belonging to w such that $Tw = 1$. Since T of nullity 0 is an automorphism, $w = 1$. This yields the familiar result:

If $w \neq 1$, *then* w *has no solution of nullity zero. (If* n *elements* $u_1, \cdots, u_n$ *generate a free group* U, *and satisfy a nontrivial relation, then* U *has rank less than* n.*)*

The question when w has solutions of nullity 1 will be settled, in the special case that w is quadratic, in Sections 4 and 5.

At the other extreme, the following observations concerning solutions of low rank are obvious.

Every w *has the trivial homomorphism as its only solution of rank* 0.

Every w *has solutions of rank* 1 (in an infinite cyclic group), *except in the case* $n = 1$ *and* $w = x_1^\alpha$, *for a nonzero integer* α.

3. QUADRATIC EQUATIONS

If w is *linear* in some x_μ, that is, given by a word $ux_\mu v$ where u and v do not contain x_μ, then it is easily seen that the single solution

$$\phi x_\nu = x_\nu \ (\nu \neq \mu), \qquad \phi x_\mu = u^{-1} v^{-1},$$

of nullity 1, constitutes a complete set of solutions.

A word is *quadratic* if each x_ν occurs in it, with exponent either +1 or -1, either twice or not at all; if a word is quadratic, the equivalent reduced word is quadratic, and we call the group element that they represent quadratic. If w is quadratic, and Sw is obtained by substituting 1 for some x_ν, then evidently Sw is also quadratic. Suppose w is quadratic, and R is an elementary regular transformation belonging to w. Then Rw results from w by replacing two parts $(x_\mu^\alpha x_\nu^\beta)^\gamma$ and x_μ^δ by $x_\mu^{\alpha\gamma}$ and $(x_\mu^\alpha x_\nu^{-\beta})^\delta$ ($\alpha, \beta, \gamma, \delta = \pm 1$), whence Rw is also quadratic. This establishes the following lemma.

LEMMA 7. *If* w *is quadratic and* T *belongs to* w, *then* Tw *is quadratic.*

PROPOSITION 8. *If* w *is quadratic, there is an effective procedure for determining the maximum rank of a solution of* w, *and for finding a solution of this maximum rank.*

Proof. If w is quadratic and T belongs to w, then w' = Tw is quadratic and of length not exceeding 2n; hence there are only finitely many such w'. For a set W of such w', let R(W) be the union of W with the set of all Rw' for w' in W and R an elementary regular transformation belonging to w'; let S(W) be the set of all Sw' for w' in W and S an elementary singular transformation belonging to w'. The ascending chain W, R(W), $R^2(W)$, $\cdots$ will become constant with a term $R^*(W)$. Starting with $W = \{w\}$, form successively the sets W, $SR^*(W)$, $(SR^*)^2(W)$, $\cdots$, $(SR^*)^n(W) = 1$. Then the maximum rank of a solution of w is n - k, where k is the smallest integer such that 1 is in $(SR^*)^k(W)$, and the process of construction of the set $(SR^*)^k(W)$ yields a solution

$$T = S_k R_{k-1,r_{k-1}} \cdots R_{k-1,1} S_{k-1} \cdots S_1 R_{0,r_0} \cdots R_{0,1}$$

of w of nullity k.

The algorithm just described can be elaborated to give a systematic enumeration of a complete set of solutions of quadratic w. The chief complication lies in the existence of cycles, that is, of nonidentical transformations T belonging to w' such that Tw' = w'. This leads to solutions containing integer parameters (telling how many times the transformation T is iterated on w'), as with the Vaught equation, where the complete solution can be given by $a = y^\alpha$, $b = y^\beta$, $c = y^{\alpha+\beta}$, for α and β integer parameters. However, consideration of the equation $a^2b^2c^2d^2 = 1$ already suggests that a complete set of solutions cannot always be given in this form by means of only a finite number of integer parameters.

The Vaught conjecture is evidently equivalent to the assertion that $w = x_1^2 x_2^2 x_3^2$ has no solutions of rank greater than 1. This is contained in the following somewhat more general theorem.

PROPOSITION 9. *A quadratic element* w *has a solution of nullity less than* 2 *if and only if some cyclic permutation of the reduced word for* w *has explicitly one of the forms*

(1) $aabccb^{-1}$, (2) $abcbac^{-1}$, (3) $abca^{-1}b^{-1}c^{-1}$.

The possibility that some of a, b, c *be empty is not excluded, and leads to the following special forms*

(4) $aabb$, (5) $abab^{-1}$, (6) $aba^{-1}b^{-1}$, (7) aa, (8) 1.

If w has a solution of nullity 0, then $w = 1$, as noted earlier. Suppose then that $w \neq 1$ and that w has a solution of nullity 1. Then $Tw = 1$ for some T belonging to w and of nullity 1, hence of the form $T = R'SR$, where R' and R are regular, R belongs to w, and S is an elementary singular transformation. This means that $w' = Rw$ reduces to 1 when one of the x_ν is replaced by 1. It follows easily, since w' is quadratic in x_ν, that w' has the form $w' = ux_\nu^{\pm 1}u^{-1}vx_\nu^{\pm 1}v^{-1}$. Since $w = R^{-1}w'$ (R an automorphism of X), it follows that w has the form $w = uvu^{\pm 1}v^{-1}$ for some u and v.

To complete the proof it will suffice to show, more generally, that if $w = uvu_1^{\pm 1}v^{-1}$ for any u, v, and for u_1 a cyclic permutation of u, then w satisfies the conclusion of the proposition. We note that the conclusion for any conjugate of w implies the same for w. We proceed first by induction on L(u). In the initial case $L(u) = 0$, we have $u = 1$ and hence $w = 1$. Assume inductively the conclusion for all $w' = u'v'u_1'^{\pm 1}v'^{-1}$ with $L(u') < L(u)$. We proceed now by a second induction on L(v).

In the initial case, $L(v) = 0$, we have $v = 1$, and $w = uu_1^{\pm 1}$. We can suppose, replacing w by a conjugate if necessary, that u is *cyclically reduced:* $L(uu) = 2L(u)$, or, explicitly, the last letter of the reduced word for u is not inverse to the first. Then we can write $u = pq$, where pq is a reduced product, and $u_1 = qp$, also reduced since $uu = pqpq$ is reduced. In the case of exponent +1, this gives $w = pqqp$. We may write $p = aba^{-1}$, $q = cdc^{-1}$, reduced products, with b, d cyclically reduced. Then $w = aba^{-1}cddc^{-1}aba^{-1}$, reduced. Setting $a^{-1}c = e$, we have $ba^{-1}wab^{-1} = bbedde^{-1}$, of form (1). In the case of exponent -1 we have $w = pqp^{-1}q^{-1}$. If $q = 1$ then $w = 1$, while otherwise this falls under the inductive hypothesis with p, q for u', v', and $L(p) < L(u)$.

For the induction on L(v), we assume the conclusion for all w' of the given form with $L(u') < L(u)$ or $L(u') = L(u)$ and $L(v') < L(v)$. If uv is not reduced, that is, unless $L(uv) = L(u) + L(v)$, we have $u = ab$, $v = b^{-1}c$, reduced, with $b \neq 1$. Then $w = acu_1^{\pm 1}c^{-1}b$ and $bwb^{-1} = bacu_1^{\pm 1}c^{-1}$, where u_1, as a permutation of $u = ab$, is a permutation of ba. Since $L(ba) = L(u)$ and $L(c) < L(v)$, the inductive hypothesis applies if $vu_1^{\pm 1}$ or $u_1^{\pm 1}v^{-1}$ is not reduced. Since the initial cases $u = 1$ or $v = 1$ have been disposed of, we are left with the case $w = uvu_1^{\pm 1}v^{-1}$, reduced. Setting $u = pq$ and $u_1 = qp$ as before, we obtain either $w = pqvqpv^{-1}$, of form (2), or

$$w = pqvp^{-1}q^{-1}v^{-1},$$

of form (3).

This completes the proof that every quadratic w with a solution of nullity less than 2 has the stated property. For the converse it suffices to consider w of form (1), (2), or (3). For w of form (1) ($w \neq 1$), the equation $abc = b$ will be linear in any letter appearing in a or c, hence will have a solution of nullity 1, which is also a solution of $w = 1$. For $w \neq 1$ of form (2) or (3), a solution of nullity 1 of the equation $ab = 1$ provides a solution of $w = 1$.

REFERENCES

1. F. Levi, *Über die Untergruppen freier Gruppen*, Math. Z. 32 (1930), 315-318.
2. ———, *Über die Untergruppen der freien Gruppen*, Math. Z. 37 (1933), 90-97.
3. J. Nielsen, *Über die Isomorphismen unendlicher Gruppen ohne Relationen*, Math. Ann. 79 (1918), 269-272.

The University of Michigan