WEDDERBURN'S LITTLE THEOREM

SHAMIL ASGARLI

Abstract. The purpose of this expository paper is to present several known proofs of Wedderburn's theorem that every finite division ring is necessarily a field.

Preface

This expository paper was the project I submitted for "Math 423: Topics in Algebra", taught by Zinovy Reichstein at the University of British Columbia in Spring 2013. I would like thank Professor Reichstein for proof-reading and making several excellent suggestions that improved the organization and the quality of the paper. I have tried to present and (at most) make tiny modifications to three existing proofs of Wedderburn's Little Theorem, and therefore I claim no original work.

1. Introduction

A division ring is a ring in which every non-zero element has a multiplicative inverse; a commutative division ring is called a field. In 1905, Wedderburn proved the remarkable theorem that every finite division ring must be a field. In other words, the finiteness of a division ring is strong enough to force commutativity. We remark that there are infinite division rings that are not fields; for instance the ring of real Hamilton quaternions is such a division ring. Many years have passed after Wedderburn's original paper (see [6]), and numerous other proofs of this theorem have been found. We will give exposition of three different proofs. We will present proofs by Ernst Witt, Hans Zassenhaus and Israel Herstein. While Witt's proof is a beautiful application of cyclotomic polynomials, Zassenhaus's proof is purely group-theoretic. Herstein's proof, on the other hand, uses assortment of results from abstract algebra, but is mainly ring-theoretic. There are other proofs of this result that are not presented in this paper; for example, Ted Kaczynski's proof (see [3]) proceeds by showing that Sylow subgroups of the multiplicative group of the finite division ring are all cyclic. See [1] for interesting history and chronological list of proofs of the *Wedderburn's Little Theorem*:

Theorem. *A finite division ring is necessarily a field.*

2. Witt's Proof

Before presenting the first proof, we need to develop some basic theory of cyclotomic polynomials. We say a complex number θ is a *primitive n-th root of unity* if $\theta^n = 1$ but $\theta^m \neq 1$ for all positive integers $m < n$. For example, $i = \sqrt{-1}$ is 4-th root of unity, while $\frac{1+\sqrt{-3}}{2}$ is a 3rd root of unity. We define n-th *cyclotomic polynomial* to be

$$\Phi_n(x) = \prod (x - \theta)$$

where the product is taken over all n-th primitives root of unity θ. For example, $\Phi_1(x) = x-1$, $\Phi_2(x) = x+1$, $\Phi_3(x) = x^2 + x + 1$, $\Phi_4(x) = x^2 + 1$. Now, the roots of the polynomial $x^n - 1$ are precisely those θ that satisfy $\theta^n = 1$, implying that θ is d-th root of unity for some $d \mid n$ (by Lagrange's Theorem). Conversely, if θ is d-th root of unity, and $d \mid n$, it is clear that $\theta^n = 1$. Thus,

$$x^n - 1 = \prod_{d|n} \Phi_d(x)$$

We show that for every $n \in \mathbb{N}$, the cyclotomic polynomial $\Phi_n(x)$ is a monic polynomial with integer coefficients. We proceed by induction. Assume the claim is true for all cyclotomic polynomials $\Phi_m(d)$ where

$m < n$. Using above relation, we can extract $\Phi_n(x)$ from the product (after all, $n \mid n$) to obtain

$$x^n - 1 = \prod_{d|n} \Phi_d(x) = \Phi_n(x) \prod_{\substack{d|n \\ d<n}} \Phi_d(x)$$

If we let

$$H(x) = \prod_{\substack{d|n \\ d<n}} \Phi_d(x)$$

then, by induction hypothesis $H(x)$ is a product of monic polynomial with integer coefficients, and consequently is a monic polynomial with integer coefficients. From the above two relations, we have $x^n - 1 = \Phi_n(x)H(x)$, and so

$$\Phi(n) = \frac{x^n - 1}{H(x)}$$

which shows that $\Phi(n)$ is a monic polynomial with integer coefficients. Next, we claim that

$$\Phi(n) \mid \frac{x^n - 1}{x^d - 1}$$

whenever $d \mid n$, $d < n$. To prove this, we observe that $x^d - 1 = \prod_{t|d} \Phi_t(x)$. But if $t \mid d$, then $t \mid n$ (because $d \mid n$). But since $n \nmid d$ (because $n > d$), $\Phi(n)$ is not a factor of $x^d - 1$. Thus, we can regroup the terms and write

$$x^n - 1 = \Phi_n(x)(x^d - 1) \prod_{\substack{k|n \\ k\nmid d \\ k\neq n}} \Phi_k(x)$$

It is now immediate that $\Phi_n(x)$ divides $\dfrac{x^n - 1}{x^d - 1}$.

Proof. Let D be a finite division ring. Then, D forms a vector space over its center $Z = \{z \in D : za = az,\ \forall a \in D\}$. Suppose Z has q elements. It follows that D must have q^n elements, where n is the dimension of the vector space D over Z. If we can show that $n = 1$, then $D = Z$, and so D is commutative (because Z is commutative).

Denote $D^\times$ to be the group of all invertible elements in D. Since D is a division ring, it follows that $D = D^\times \cup \{0\}$. Thus, $|D^\times| = q^n - 1$. For each $a \in D$, we let $C(a) = \{x \in D : ax = xa\}$ be the centralizer of a in D. It is immediate that $Z \subset C(a)$, so $N(a)$ forms a vector space over Z, meaning that $C(a)$ has q^d elements for some $d \in \mathbb{N}$ (where r depends on a). We note that $C(a)^\times$ contains $q^d - 1$ elements, and $C(a)^\times$ is a subgroup of $D^\times$. From Lagrange's theorem we get $q^d - 1 \mid q^n - 1$. We claim that $d \mid n$. Indeed, using division algorithm we can write $n = Qd + r$ where $0 \leq r < d$. Since $q^d - 1 \mid q^n - 1$, we must have

$$q^d - 1 \mid \left[(q^n - 1) - (q^{n-d} + q^{n-2d} + \cdots q^{n-Qd})(q^d - 1)\right] = q^{n-Qd} - 1 = q^r - 1$$

We have shown $q^d - 1 \mid q^r - 1$. But $0 \leq r < d$. If $r > 0$, then $q^d - 1 > q^r - 1$, and this contradicts $q^d - 1 \mid q^r - 1$. Thus, $r = 0$ and so $d \mid n$. The center of the $D^\times$ is simply $Z^\times$. Applying the class equation to group $D^\times$ we obtain

$$|D^\times| = |Z^\times| + \sum_{a \notin Z} \frac{|D^\times|}{|N(a)|}$$

which translates into

$$q^n - 1 = q - 1 + \sum_{\substack{d|n \\ d<n}} \frac{q^n - 1}{q^d - 1}$$

From our above discussion about cyclotomic polynomials, we know that the n-th cyclotomic polynomial $\Phi_n(q)$ divides $\dfrac{q^n - 1}{q^d - 1}$ for each $d \mid n$ with $d < n$. But also $\Phi_n(q)$ divides $q^n - 1$. As a result of the class equation above, $\Phi_n(q)$ divides $q - 1$. However, if $n > 1$ then $|q - \theta| > |q - 1|$ for any θ that is a primitive n-th root of unity. Thus,

$$\Phi_n(q) = \prod (q - \theta) > q - 1$$

In particular, $\Phi(n)$ cannot divide $q-1$. We have reached a contradiction, and this concludes the proof of Wedderburn's Theorem.

□

3. Zassenhaus's proof

We start by examining some facts from Galois Theory. See [4] (Chapter 3) for more detailed account. Let F be a finite field, and $S \subset F$ be its subfield. Given $x \in F$, we define *norm* of x over S (denoted by $N_{F/S}(x)$) to be the product of all conjugates of x in $G_{F/S}$, where $G_{F/S}$ is the group of automorphisms of F that leave S invariant.

$$N_{F/S}(x) = \prod_{\sigma \in G_{F/S}} \sigma(x)$$

Suppose F is a finite field with a characteristic $p > 0$. It is a well-known fact that only field automorphisms F are of the form σ^k where $\sigma : F \to F$ is defined by $\sigma(a) = a^p$ for each $a \in F$. The automorphism σ is usually referred to as Frobenius automorphism. Now suppose $S \subset F$ is a subfield with q elements, and F has q^n elements. It follows that $G_{F/S}$ consists of automorphisms of the form $x \mapsto x^{q^k}$ where $0 \leq k \leq n-1$. Thus

$$N_{F/S}(x) = \prod_{\sigma \in G_{F/S}} \sigma(x) = x^{1+q+q^2+\cdots+q^{n-1}} = x^{\frac{q^n-1}{q-1}}$$

Proposition. Suppose $S \subset F$ is a subfield with q elements, where F has q^n elements. Then, each element of S can be written as norm of some element of F.

Proof. Consider the map $\Psi : F^\times \to S^\times$ defined by $\Psi(x) = N_{F/S}(x)$, which is easily checked to be a group homomorphism. We have

$$\ker(\Psi) = \{a \in F^\times : x^{\frac{q^n-1}{q-1}} = 1\}$$

Since F is a field, the equation $x^{\frac{q^n-1}{q-1}} = 1$ has at most $\frac{q^n-1}{q-1}$ solutions; hence $\ker(\Psi)$ has at most $\frac{q^n-1}{q-1}$ elements. By first isomorphism theorem, $\mathrm{im}(\Psi)$ is isomorphic to $F^\times/\ker(\Psi)$, and so $\mathrm{im}(\Psi)$ has at least

$$\frac{|F^\times|}{|\ker(\Psi)|} = \frac{q^n-1}{\frac{q^n-1}{q-1}} = q-1$$

elements. But then $q-1 \leq |\mathrm{im}(\Psi)| \leq |S^\times| = q-1$. It follows that $\mathrm{im}(\Psi) = S^\times$, and the map Ψ is onto. Consequently, every element of S can be realized as norm of some element of F. □

We recall from group theory that given a group G, and a subgroup $H \subset G$, the *normalizer* of H in G is defined by $N_G(H) = \{a \in G : aH = Ha\}$ and the *centralizer* of H in G is defined by $C_G(H) = \{a \in G : ah = ha \ \forall h \in H\}$. We will first prove a sufficient condition for a finite group to be abelian, and then show that the group of units in any finite division ring satisfies this group-theoretic property.

We will closely follow Zassenhaus's original paper (see [5]).

Theorem. *Let G be a finite group. If $N_G(H) = C_G(H)$ for all abelian subgroups $H \subset G$, then G is abelian.*

Proof. We proceed by induction on the number of elements in G. Let G be a finite group of order n with the property that $N_G(H) = C_G(H)$ for all abelian subgroups $H \subset G$. We suppose that the theorem holds for all finite groups of order strictly less than n. Then, every proper subgroup of G has order less than n, and satisfies the condition on the equality of normalizer and centralizer for abelian subroups, and so by induction hypothesis it is abelian. Thus, every proper subgroup of G is abelian. Now, we divide the proof into two cases, whether or not the center of the group $Z = \{z \in G : ag = ga \ \forall g \in G\}$ is trivial.

Case 1: $|Z| > 1$, i.e. the center of G is not trivial.
Then the quotient group G/Z has order less than n. We will show that G/Z is an abelian group. Assume $\overline{H}$ is any abelian subgroup of G/Z. We may assume $\overline{H}$ is a proper subgroup of G/Z, for otherwise $G/Z = \overline{H}$ is abelian, as desired. By correspondence theorem, there exists a subgroup $H \subset G$ containing Z such that $H/Z = \overline{H}$. It is clear that H is a proper subgroup of G (because otherwise if $H = G$, then $\overline{H} = H/Z = G/Z$, but $\overline{H}$ is a proper subgroup of G/Z, contradiction). As we observed in above paragraph, properness of H in G together with induction hypothesis forces H to be abelian. We will now prove that $N_{G/Z}(\overline{H}) = C_{G/Z}(\overline{H})$.

The inclusion $C_{G/Z}(\overline{H}) \subset N_{G/Z}(\overline{H})$ is trivially true. For the reverse inclusion, let $\overline{X}$ be any element of $N_{G/Z}(\overline{H})$. Then, $(\overline{X})\overline{H}(\overline{X})^{-1} = \overline{H}$. If $\pi : G \to G/Z$ is the natural projection map, then for every $x \in \pi^{-1}(\overline{X})$, we have $xHx^{-1} = H$. So x lies in the normalizer of H, and since H is abelian group, x lies in the centralizer of H. But then, $xhx^{-1} = h$ for all $h \in H$. Applying the natural projection map π, we get $\overline{x}\overline{h}\overline{x}^{-1} = \overline{h}$. This is true for all $x \in \pi^{-1}(\overline{X})$. Consequently, x lies in the centralizer $C_{G/Z}(\overline{H})$. We deduce that $N_{G/Z}(\overline{H}) = C_{G/Z}(\overline{H})$. As H was arbitrary abelian subgroup of G/Z, and $|G/Z| < n$ by induction hypothesis it follows that G/Z is abelian, as desired. We will now show that G must also be abelian. Let $a, b \in G$ be any two arbitrary elements. By definition of center, $bz = zb$ for each $z \in Z$. Furthermore, Z itself is clearly an abelian group; hence the subgroup generated by $(b, Z) = (b, Z)$ is an abelian subgroup of G. Since G/Z is abelian, we have

$$(aba^{-1}Z) = (aZ)(bZ)(a^{-1}Z) = (bZ)(aZ)(a^{-1}Z) = bZ(aa^{-1}Z) = bZ$$

So $aba^{-1}Z = bZ$. Thus, there exists $z \in Z$ with $aba^{-1} = bz$. Since $za = az$ for all $z \in Z$, it is clear that if $y \in \langle b, Z\rangle$, then $aya^{-1} \in \langle b, Z\rangle$. Consequently, $a\langle b, Z\rangle a^{-1} = \langle b, Z\rangle$, meaning that a is the normalizer of $\langle b, Z\rangle$, and so a is in the centralizer of $\langle b, Z\rangle$ (because $\langle b, Z\rangle$ is an abelian subgroup of G). In particular, a commutes with $b \in \langle b, Z\rangle$, giving us $ab = ba$. As $a, b \in G$ were arbitrary, we deduce that G is abelian.

Case 2: $|Z| = 1$, i.e. the center of G is trivial.
Assume M_1 and M_2 are two distinct maximal proper subgroups of G. By induction hypothesis, M_1 and M_2 are abelian. Also, the subgroup generated by M_1 and M_2 must be the entire group, i.e. $\langle M_1, M_2\rangle = G$ (otherwise, maximality of M_1 would be contradicted). Now, if $x \in M_1 \cap M_2$, then x commutes with every element of M_1 and x commutes with every element of M_2. Since M_1 and M_2 are abelian, it follows that x commutes with every element of $\langle M_1, M_2\rangle = G$. As a result, $x \in Z = \{1\}$. We conclude that any two distinct maximal subgroups of G intersect at the identity element. Now suppose M is any maximal proper subgroup of G. We will now analyze conjugate subgroups of M. First, for any $x \in G$, the conjugate subgroup xMx^{-1} is maximal (Indeed, if $xMx^{-1} \subsetneq Q$ then $M \subsetneq x^{-1}Qx$, and so $x^{-1}Qx = G$, implying that $Q = G$). Suppose $xMx^{-1} = yMy^{-1}$ for some $x, y \in G$. Then, $(y^{-1}x)M(y^{-1}x)^{-1} = M$, and so $y^{-1}x$ is in the normalizer of M, and hence, in the centralizer of M (because M is an abelian subgroup of G). So $y^{-1}x$ commutes with every element of M, and consequently commutes with every element of $\langle y^{-1}x, M\rangle$. If $y^{-1}x \notin M$, then by maximality of M, $\langle y^{-1}x, M\rangle = G$, and so $y^{-1}x$ commutes with every element of G, i.e. $y^{-1}x \in Z$, so $y^{-1}x$ is the identity element (as we are in Case 2), and so $y^{-1}x \in M$ contradiction, as we were assuming $y^{-1}x \notin M$. Thus, whenever $xMx^{-1} = yMy^{-1}$, it follows that $y^{-1}x \in M$. The converse is also true: if $y^{-1}x \in M$, then $(y^{-1}x)M(x^{-1}y) = M$, and so $yMy^{-1} = y(y^{-1}xMx^{-1}y)y^{-1} = xMx^{-1}$. So, $xMx^{-1} = yMy^{-1}$ if and only if $y^{-1}x \in M$. Equivalently, $xMx^{-1} = yMy^{-1}$ if and only if $xM = yM$. We deduce that the number of conjugates of M is equal to the number of left cosets of M, i.e. there are $|G|/|M|$ conjugates of M. Since any two distint maximal subgroups of G intersect at identity (as we proved above), and since each conjugate subgroup contains $|M| - 1$ non-identity elements, the total number of non-identity elements in all conjugates of M is

$$\frac{|G|}{|M|}(|M| - 1) = |G| - \frac{|G|}{|M|} \tag{1}$$

We now have enough tools to show G is abelian. Take a non-identity element $x \in G$. If $(x) = G$, then G is cyclic, and hence abelian. We may assume $(x) \subsetneq G$, so that (x) is contained in some maximal subgroup $M \subset G$. Since $|M| > 1$ (because $(x) \subset M$ and $|(x)| > 1$), we see from equation (1) that M and its conjugates supply $|G|/2$ non-identity elements of G. If G contained any element w that is not in M or in any of the conjugates, then w would be contained in some other maximal subgroup, say $P \subset G$. Since any two distinct maximal subgroups of G are distinct, P and its conjugates supply $|G|/2$ non-identity elements of G, all of which all of which are different from the above $|G|/2$ non-identity elements supplied by M and its conjugates. Thus, P, M together with conjugates supply $|G|/2 + |G|/2 = |G|$ non-identity elements of G. Adding the identity element, we get a total of $|G| + 1$ elements in G, contradiction. Hence, G cannot contain any element w that is not in M or any of its conjugates. We conclude that M and its conjugates supply all the elements of G. From equation (1), this means

$$|G| - \frac{|G|}{|M|} + 1 = |G|$$

giving us $\frac{|G|}{|M|} = 1$, so that $G = M$. This contradicts the fact that M is a proper subgroup of G, and establishes the theorem. $\square$

We will now show that the above group-theoretic property is satisfied by finite division rings.

Theorem. *Suppose D is a division ring. If G is any abelian subgroup of $D^\times$, the normalizer of G coincides with the centralizer of G.*

Proof. Let G be any abelian subgroup of $D^\times$. Denote $N(G)$ and $C(G)$ to be the normalizer and centralizer of group G, respectively. It is always true that $C(G) \subset N(G)$. We want to show the reverse inclusion. Suppose $x \in D^\times$ such that $xGx^{-1} = G$, i.e. $x \in N(G)$. Our goal is to show that $xax^{-1} = a$ for all $a \in G$, i.e. $x \in C(G)$. Since $x^{|D^\times|} = 1$, and $1 \in C(G)$, there exists a smallest positive integer m such that $x^m \in C(G)$. By definition, for all $a \in G$, we have

$$x^m a x^{-m} = a$$

and for each positive integer k satisfying $0 < k < m$, there is an $a_k \in G$ such that

$$x^k a_k x^{-k} \neq a_k$$

Let H be the subgroup of $D^\times$ generated by x^m and G. Because $x^m a = a x^m$ for all $a \in G$, and G is abelian, it follows that H is abelian. As $xGx^{-1} = G$, and $xx^m x^{-1} = x^m$, it follows that

$$xHx^{-1} = H$$

Define a new set F, which consists of all elements in D that can be written as a finite sum of elements in H. In other words,

$$F = \left\{ \sum_{i=1}^{t} b_i \ : \ b_i \in H, t \in \mathbb{N} \right\}$$

We claim that F is a field. It is easy to see that F is closed under addition, and multiplication. Therefore, $F^\times$ is a multiplicative subgroup of $D^\times$. Since H is abelian, F is commutative ring, showing that F is a field.

We have shown that $xHx^{-1} = H$. Using $xHx^{-1} \subset H$, it follows that if $y \in F$, then $xyx^{-1} \in F$ (simply because elements of F are finite sums formed from elements of H). Similarly, using $H \subset xHx^{-1}$, it follows that every element of F can be written of the form xyx^{-1} with $y \in F$. Also,

$$x(a+b)x^{-1} = xax^{-1} + xbx^{-1}$$
$$x(ab)x^{-1} = (xax^{-1})(xbx^{-1})$$

As a result, we have an automorphism of the field F, given by map $a \to xax^{-1}$ for each $a \in F$. We will denote this map by σ. So $\sigma(a) = xax^{-1}$. Recall that we chose $m \in \mathbb{N}$ to be the smallest positive integer such that $x^m \in C(G)$. We want to show that $m = 1$. Assume, to the contrary, that $m > 1$. Then, by easy induction, we have

$$\sigma^m(a) = x^m a x^{-m} = a$$

for all $a \in F$. But if $0 < k < m$, there exists $a_k \in G \subset H \subset F$ such that

$$\sigma^k(a) = x^k a_k x^{-k} \neq a_k$$

We conclude that σ^m is the identity automorphism, while the automorphism σ^k for $0 < k < m$ is not the identity.

Now $\sigma(x^m) = xx^m x^{-1} = x^m$, showing that x^m is invariant under σ. Let S denote the subfield of F, consisting of all those elements in F that are invariant under σ. In particular, $x^m \in S$. Then, by **Proposition** proved in the beginning, x^m is the norm $N_{F/S}(y)$ of some element $y \in F$. Hence,

$$x^m = y\sigma(y)\sigma^2(y)\cdots\sigma^{m-1}(y)$$

We consider the expression

$$f(x,y) = (x-y)(1 + y^{-1}x + y^{-1}\sigma^{-1}(y)x^2 + \cdots + y^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1})$$

After expanding we obtain

$$f(x,y) = x + xy^{-1}x + xy^{-1}\sigma^{-1}(y)x^2 + \cdots + xy^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1}$$

$$- y - x - \sigma^{-1}(y)x^2 - \cdots - \sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1}$$

In order to simplify terms in the first row, we observe that for each $1 \leq r \leq m$,

$$\begin{aligned} & xy^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(r-2)}(y)x^{r-1} \\ &= (xy^{-1}x^{-1})(x\sigma^{-1}(y)x^{-1})\cdots(x\sigma^{-(r-2)}(y)x^{-1})x^r \\ &= \sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(r-1)}(y)x^r \end{aligned}$$

Returning to previous equation, and substituting the above expression for each $1 \leq r \leq m$ we get lots of cancellations:

$$\begin{aligned} f(x,y) &= x + xy^{-1}x + xy^{-1}\sigma^{-1}(y)x^2 + \cdots + xy^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1} \\ &\quad - y - x - \sigma^{-1}(y)x^2 - \cdots - \sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1} \\ &= -y + \sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-1)}(y)x^m \\ &= -y + y = 0 \end{aligned}$$

where in the next to last step, we used $x^m = y\sigma(y)\sigma^2(y)\cdots\sigma^{m-1}(y)$. We have shown that $f(x,y) = 0$, that is,

$$(x-y)(1 + y^{-1}x + y^{-1}\sigma^{-1}(y)x^2 + \cdots + y^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1}) = 0$$

Since D is an integral domain, one of the factors in the above inequality must be zero, i.e. $1 + y^{-1}x + y^{-1}\sigma^{-1}(y)x^2 + \cdots + y^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1} = 0$ or $x = y$. If $x = y$, then $x \in F$ (because $y \in F$), and so x commutes with every element in $F \subset G$, and so $xax^{-1} = a$ for all $a \in G$, contradicting our assumption that $m > 1$. So it must be the case that

$$1 + y^{-1}x + y^{-1}\sigma^{-1}(y)x^2 + \cdots + y^{-1}\sigma^{-1}(y)\sigma^{-2}(y)\cdots\sigma^{-(m-2)}(y)x^{m-1} = 0$$

Since the above equation holds, we know that the equation

$$1 + \sum_{\delta=1}^{t} c_{j_\delta}x^{j_\delta} = 0 \tag{2}$$

holds, where $0 < j_1 < j_2 < \cdots < j_t < m$ and $0 \neq c_{j_\delta} \in F$ for each $1 \leq \delta \leq t$. Among the relations of the form above, there exists one with minimal $t \in \mathbb{N}$. Suppose the above equation is (2) one of them. Recall that, for each $0 < k < m$, there exists $a_k \in G$ such that

$$\sigma^k(a) = x^k a_k x^{-k} \neq a_k$$

Since $0 < j_1 < m$, such $a_{j_1} \in G$ exists. Multiply both sides of equation (2) on the left by $a_{j_1}^{-1}$, and on the right but a_{j_1}, we obtain

$$1 + \sum_{\delta=1}^{t} a_{j_1}^{-1}c_{j_\delta}x^{j_\delta}a_{j_1} = 0 \tag{3}$$

Let $d_{j_\delta} = a_{j_1}^{-1}c_{j_\delta}(x^{j_\delta}a_{j_1}x^{-j_\delta})$. The equation (3) becomes:

$$\begin{aligned} 0 &= 1 + \sum_{\delta=1}^{t} a_{j_1}^{-1}c_{j_\delta}x^{j_\delta}a_{j_1} \\ &= 1 + \sum_{\delta=1}^{t} a_{j_1}^{-1}c_{j_\delta}(x^{j_\delta}a_{j_1}x^{-j_\delta})x^{j_\delta} \\ &= 1 + \sum_{\delta=1}^{t} d_{j_\delta}x^{j_\delta} \end{aligned}$$

Recall that σ defined by $\sigma(a) = xax^{-1}$ is an automorphism on F. And so σ^{j_δ} leaves F invariant; in particular $d_{j_\delta} \in F$ for each $1 \leq \delta \leq t$. We have

$$1 + \sum_{\delta=1}^{t} c_{j_\delta} x^{j_\delta} = 0 \tag{4}$$

$$1 + \sum_{\delta=1}^{t} d_{j_\delta} x^{j_\delta} = 0 \tag{5}$$

Subtract equation (5) from equation (4) to obtain

$$(c_{j_1} - d_{j_1})x^{j_\delta} + \sum_{\delta=2}^{t} (c_{j_\delta} - d_{j_\delta})x^{j_\delta} = 0$$

Multiplying both sides by x^{-j_1} on the right, we have

$$c_{j_1} - d_{j_1} + \sum_{\delta=2}^{t} (c_{j_\delta} - d_{j_\delta})x^{j_\delta - j_1} = 0$$

Using the definition of d_{j_v} and noting that $c_{j_1} \in F$, we observe

$$d_{j_1} = a_{j_1}^{-1} c_{j_1} (x^{j_1} a_{j_1} x^{-j_1}) \neq a_{j_1}^{-1} c_{j_1} a_{j_1} = c_{j_1}$$

We have $d_{j_1} \neq c_{j_1}$. So $0 \neq c_{j_1} - d_{j_1} \in F$ has a multiplicative inverse. Multiplying equation (5) by $(c_{j_1} - d_{j_1})^{-1}$ on the left, we get

$$1 + \sum_{\delta=2}^{t} p_{j_\delta} x^{j_\delta - j_1} = 0 \tag{6}$$

where $p_{j_\delta} = (c_{j_1} - d_{j_1})^{-1}(c_{j_\delta} - d_{j_\delta}) \in F$ (as $c_{j_\delta}, d_{j_\delta} \in F$ for each j_δ). But this is a contradiction, because we choose t to be minimal among relations of the form above; and the equation (6) has $t - 1$ terms in the sum. Thus, the assumption that $m > 1$ leads to a contradiction. We must have $m = 1$, and $xax^{-1} = x$ for each $a \in G$. In other words, $x \in C(G)$, and so $N(G) \subset C(G)$. We conclude that $N(G) = C(G)$ holds for every abelian subgroup G of a finite division ring D. □

The above two theorems together imply that $D^\times = D \setminus \{0\}$ must be abelian. In other words, D is commutative, and Wedderburn's Little Theorem has been proved.

4. Herstein's proof

We proceed by proving series of lemmas. Our first lemma will be used in the very last step of the proof. We include it in the beginning, so as not to interfere with the intertwined lemmas that follow afterwards.

Lemma 1. Suppose F is a finite field, and $\alpha, \beta \in F$ satisfying $\alpha \neq 0$, $\beta \neq 0$. Then, we can find two elements $a, b \in F$ such that $1 + \alpha a^2 + \beta b^2 = 0$.

Proof. If characteristic of F is 2 (i.e. $2x = 0 \forall x \in F$), then F has order 2^n for some $n \geq N$. Thus, every element $x \in F$ satisfies $x^{2^n} = x$. In particular, every element is a square. So $\alpha = a^2$ for some $a \in F$. If we use this a, and $b = 0$, we obtain that

$$1 + \alpha a^2 + \beta b^2 = 1 + (\alpha)(\alpha)^{-1} + 0 = 1 + 1 = 0$$

as desired.

Now suppose F has characteristic p, where p is some odd prime. Then, F has p^n elements for some $n \in \mathbb{N}$. Let $W_\alpha = \{1 + \alpha x^2 : x \in F\}$. We can count how many elements there are in W_α. If $1 + \alpha x^2 = 1 + \alpha y^2$, then we get $x^2 = y^2$, which is possible only when $x = y$ or $x = -y$. So if $x \neq 0$, the terms x and $-x$ gives only one contribution to the set W_α. But $x = 0$ gives only one contribution (since $-0 = 0$). Thus, the number of elements in $W_\alpha = 1 + (p^n - 1)/2 = (p^n + 1)/2$. Similarly, if we define $W_\beta = \{-\beta x^2 : x \in F\}$, the exact same reasoning tells us that W_β has also $(p^n + 1)/2$ elements. But now,

$$|W_\alpha| + |W_\beta| = \frac{p^n + 1}{2} + \frac{p^n + 1}{2} = p^n + 1 > p^n = |F|$$

By pigeonhole principle, there exists a common element $g \in W_\alpha \cup W_\beta$. By definition of W_α and W_β, we have $g = 1 + \alpha a^2$ for some $a \in F$, and $g = -\beta b^2$ for some $b \in F$. Equating these two expressions, we arrive at $1 + \alpha a^2 = -\beta b^2$, or equivalently $1 + \alpha a^2 + \beta b^2 = 0$, as desired.

□

Let R be a ring. For each $a \in R$, define $\psi_a : R \to R$ by $\psi_a(x) = xa - ax$. Our first lemma is concerned about the powers of ψ_a, which in this context means composition of ψ_a by itself. We will write ψ_a^m to mean m-fold composition of ψ_a, i.e $\psi_a^m = \psi \circ \psi \circ \cdots \circ \psi$, where composition is nested m times.

Lemma 2. The function $\psi_a(x) = xa - ax$ defined as above satisfies the following identity:

$$\psi_a^m(x) = \sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m-j}$$

for every $m \in \mathbb{N}$.

Proof. We proceed by induction. When $m = 1$, we get $\psi_a^m(x) = \psi(x) = xa - ax$ which agrees with the above identity. Suppose the claim is proved for some $m \in N$. Then,

$$\begin{aligned}
\psi_a^{m+1}(x) = \psi_a(\psi_a^m)(x) &= \psi\left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m-j}\right) \\
&= \left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m-j}\right) a - a\left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m-j}\right) \\
&= \left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m+1-j}\right) + \left(\sum_{j=0}^{m} (-1)^{j+1} \binom{m}{j} a^{j+1} x a^{m-j}\right) \\
&= \left(\sum_{j=0}^{m} (-1)^j \binom{m}{j} a^j x a^{m+1-j}\right) + \left(\sum_{j=1}^{m+1} (-1)^j \binom{m}{j-1} a^j x a^{m-(j-1)}\right) \\
&= xa^{m+1} + \left(\sum_{j=1}^{m} (-1)^j \left[\binom{m}{j} + \binom{m}{j-1}\right] a^j x a^{m+1-j}\right) + (-1)^{m+1} a^{m+1} x \\
&= xa^{m+1} + \left(\sum_{j=1}^{m} (-1)^j \binom{m+1}{j} a^j x a^{m+1-j}\right) + (-1)^{m+1} a^{m+1} x \\
&= \sum_{j=0}^{m+1} (-1)^j \binom{m+1}{j} a^j x a^{m+1-j}
\end{aligned}$$

Note that in next to final step we used the formula:

$$\binom{m+1}{j} = \binom{m}{j} + \binom{m}{j-1}$$

which can be easily verified. It is also immediate from Pascal's triangle. So we have proved the claim for ψ_a^{m+1}. By the principle of mathematical induction, the lemma is proved. □

Corollary. Suppose R is a ring of characteristic p, where p is a prime (i.e. $px = 0$ for all $x \in R$). Then $\psi_a^{p^m}(x) = xa^{p^m} - a^{p^m} x$.

Proof. If $p = 2$, then $\psi_a^2(x) = xa^2 - 2axa - a^2 x = xa^2 - a^2 x$. In other words, $\psi_a^2(x) = \psi_{a^2}(x)$. Suppose that $\psi_a^{2^m}(x) = xa^{2^m} - a^{2^m} x$ for some $m \geq 1$. Then,

$$\psi_a^{2^{m+1}}(x) = (\psi_a^2)^{2^m}(x) = (\psi_{a^2})^{2^m}(x) = x(a^2)^{2^m} - (a^2)^{2^m} x = xa^{2^{m+1}} - a^{2^{m+1}} x$$

proving the claim for $m + 1$. By induction, the proof is complete when $p = 2$.

Similarly, assume p is an odd prime. Using the lemma above, and the fact that p divides the binomial coefficient $\binom{p}{i}$ for $1 \leq i \leq p-1$, we obtain

$$\psi_a^p(x) = xa^p - a^p x$$

In other words $\psi_a^p(x) = \psi_{a^p}(x)$. Suppose $\psi_a^{p^m}(x) = xa^{p^m} - a^{p^m}x$ for some $m \geq 1$. Then, we obtain that

$$\psi_a^{p^{m+1}}(x) = (\psi_a^p)^{p^m}(x) = (\psi_{a^p})^{p^m}(x) = x(a^p)^{p^m} - (a^p)^{2^p}x = xa^{p^{m+1}} - a^{p^{m+1}}x$$

which proves the claim for $m+1$. The proof is complete by induction. □

Proposition: Suppose D is a division ring of characteristic $p > 0$ with centre Z, and let $P = \{0, 1, 2, ..., p-1\} \subset \mathbb{Z}$ be its prime subfield. If $a \in D$ but $a \notin \mathbb{Z}$, then there exists $x \in D$ such that
1) $xax^{-1} \in P(a)$
2) $xax^{-1} \neq a$
3) $xax^{-1} = a^i$ for some $i \geq 2$
Here $P(a)$ stands for the field obtained by adjoining a to the field P.

Proof. Assume $a \in D$ but $a \notin \mathbb{Z}$. We consider the map $\psi_a : D \to D$ defined by $\psi_a(x) = xa - ax$ some of whose properties we have already established above. Since $P(a)$ is a finite field of characteristic $p > 0$, it has p^m elements for some $m \geq 1$. Thus, $u^{p^m} = u$ for each $u \in P(a)$. In particular, since $a \in P(a)$, we have $a^{p^m} = a$. From the corollary above, we have that $\psi_a^{p^m}(x) = xa^{p^m} - a^{p^m}x = xa - ax = \psi_a(x)$. Hence, $\psi_a^{p^m} = \psi_a$.

For each $\lambda \in P(a)$, we have $\lambda a = a\lambda$ because elements of $P(a)$ are simply polynomials in a, which clearly commute with the element a. But now

$$\psi_a(\lambda x) = (\lambda x)a - a(\lambda)x = \lambda(xa) - \lambda(ax) = \lambda(xa - ax) = \lambda(\psi_a(x))$$

Since this is true for every $x \in D$, we deduce that the maps ψ_a and $\lambda I : D \to D$ defined by $\lambda I(y) = \lambda y$ commute i.e. $\psi_a(\lambda I) = (\lambda I)\psi_a$ for every $\lambda \in P(a)$. Since $P(a)$ is a field with p^m elements, we have the following factorization

$$u^{p^m} - u = \prod_{\lambda \in P(a)} (u - \lambda)$$

Since ψ_a and λI commute for each $\lambda \in P(a)$, we can formally substitute the function ψ_a instead of x, and the above equation would still hold (we replace λ by its operator λI). Also, we use $\psi_a^{p^m} = \psi_a$ to get

$$0 = \psi_a^{p^m} - \psi_a = \prod_{\lambda \in P(a)} (\psi_a - \lambda I)$$

We can rewrite the above factorization as

$$0 = \psi_a(\psi_a - \lambda_1 I)(\psi_a - \lambda_2) \cdots (\psi_a - \lambda_r)$$

where $r = p^m - 1$, and $\lambda_i \neq 0$ for each $1 \leq i \leq r$. If $(\psi_a - \lambda_i)(x) \neq 0$ for every $x \in D$, and for every $1 \leq 1 \leq r$, then the above equation would imply $\psi_a(x) = 0$ for every $x \in D$, and so $xa - ax = 0$, giving us $xa = ax$ for each $x \in D$. But then, $a \in Z$, contradiction. Thus, there must exist some $x \in D$ with $x \neq 0$ and $1 \leq j \leq r$ such that $(\psi_a - \lambda_j I)x = 0$, meaning that $xa - ax - \lambda_j x = 0$, or equivalently, $xa = ax + \lambda_j x$. Multiplying both sides by x^{-1} (which is allowed, since D is a division ring, and $x \neq 0$), we obtain $xax^{-1} = a + \lambda_j \in P(a)$ as $a, \lambda_j \in P(a)$, proving statement (1). Since $\lambda_j \neq 0$, we must have $xax^{-1} \neq a$. This proves the statement (2). Now, suppose $a \in P(a)$ has order s. Then, all the roots of $y^s - 1$ in the field $P(a)$ are $1, a, a^2, ..., a^{s-1}$ since all of these are distinct. Since $(xax^{-1})^s = xa^s x^{-1} = xx^{-1} = 1$, we conclude that $xax^{-1} \in P(a)$ is a root of the polynomial $y^s - 1 = 0$, meaning that $xax^{-1} = a^i$ for some $i \geq 2$ (note that $i \neq 1$, because $xax^{-1} \neq a$). This proves statement (3), and completes the proof of the proposition. □

We will now prove Wedderburn's Theorem.

Proof. Let D be a finite division ring. We want to show that D is commutative. We proceed by induction on the number of elements. In other words, we assume that Wedderburn's theorem holds for all finite division rings of order less than $|D|$. Suppose Z is centre of D.

Claim 1. Suppose $a, b \in D$ such that $b^t a = ab^t$ for some $t \in \mathbb{N}$ and $ab \neq ba$. Then $b^t \in D$.

Proof of Claim 1. Consider the centralizer $C(b^t) = \{y \in D : b^t y = y b^t\}$. Note that $C(b^t)$ is a subdivision ring of D. If $C(b^t) \neq D$, then $C(b^t)$ is a proper subdivision ring of D, and so by induction hypothesis we see that $C(b^t)$ must be commutative; since $a \in C(b^t)$ and $b \in C(b^t)$, this means $ab = ba$, contradiction. Thus, it must be the case that $C(b^t) = D$, and hence $b^t \in Z$. □

Given $w \in D$, we let $m(w)$ to be the smallest positive integer such that $w^{m(w} \in Z$. Since w has a finite order, $m(w)$ is well-defined. We call $m(w)$ *the order of* w *relative* to Z.

Claim 2. Let $a \in D$ with $a \in Z$ be an element with the smallest order r relative to Z. Then r must be a prime number.

Proof of Claim 2. Assume, to the contrary, that r is composite. Then, $r = r_1 r_2$ where $r_1 < r$ and $r_2 < r$. Let $a_1 = a^{r_1} \in D$. By minimality of r, $a_1 \notin Z$. However, $a_1^{r_2} = a^r = 1$. We deduce that the order of a_1 relative to Z is at most r_2. Since $r_2 < r$, this contradicts the minimality of r. □

We assume, to the contrary, that D is not a field. Then, as in the statement of **Claim 2**, we can find $a \in D$ with $a \notin Z$ such that order of a relative to Z is minimal. Suppose order of a relative to Z is r. Then we have shown that r must be prime. By applying **Proposition** above, there exists $x \in D$ such that $xax^{-1} = a^i$ and $xa \neq ax$. Now,

$$x^2 a x^{-2} = x(xax^{-1})x^{-1} = x(a^i)x^{-1} = (xax^{-1})^i = (a^i)^i = a^{i^2}$$

Similarly,

$$x^3 a x^{-3} = x(x^2 a x^{-2})x^{-1} = x(a^{i^2})x^{-1} = (xax^{-1})^{i^2} = (a^i)^{i^2} = a^{i^3}$$

It is easy to see inductively that $x^t a x^{-t} = a^{i^t}$ for each $t \in \mathbb{N}$. In particular, $x^{r-1} a x^{-(r-1)} = a^{i^{r-1}}$.

Claim 3. r does not divide i.

Proof of Claim 3. Indeed, if $r \mid i$, then $i = sr$ for some $s \in N$, and from $xax^{-1} = a^i$, we would get $xa = (a^r)^s x$. Since $a^r \in Z$, we get $(a^r)^s \in Z$ (as Z is a subring), and so $xa = (a^r)^s x = x(a^r)^s$. Cancelling x from both sides, we arrive at $a = (a^r)^s$ which gives $a \in Z$, contradiction. We deduce that $r \nmid i$. □

Since r is prime and $r \nmid i$, we have $i^{r-1} \cong 1 \pmod{p}$ by Fermat's Little Theorem. And so $i^{r-1} = 1 + cr$ for some integer c. Let $\lambda = a^{cr} = (a^r)^c \in Z$. Then,

$$a^{i^{r-1}} = a^{1+cr} = a a^{cr} = a\lambda = \lambda a$$

The equation $x^{r-1} a x^{-(r-1)} = a^{i^r}$ gives $x^{r-1} a = a^{i^r} x^{r-1} = \lambda a x^{r-1}$. So $x^{r-1} a = \lambda a x^{r-1}$.

Claim 4. We must have $\lambda \neq 1$.

Proof of Claim 4. Assume, to the contrary, that $\lambda = 1$. Then, $x^{r-1} a = \lambda a x^{r-1}$ becomes $x^{r-1} a = a x^{r-1}$. We still have $xa \neq ax$. By Claim 1, we obtain $x^{r-1} \in Z$. This contradicts minimality of r, as $r$is by definition the smallest positive integer with $a^r \in Z$. Hence, $\lambda \neq 1$. □

Let $b = x^{r-1}$. The equation $x^{r-1} a = \lambda a x^{r-1}$ becomes $ba = \lambda ab$, that is $bab^{-1} = \lambda a$. Since $\lambda \in Z$, we have

$$\lambda^r a^r = (\lambda a)^r = (bab^{-1})^r = ba^r b^{-1} = a^r b b^{-1} = a^r$$

where in the last step we used $a^r \in Z$. So $\lambda^r a^r = a^r$, which implies $\lambda^r = 1$.

Claim 5. If $y \in D$ and $y^r = 1$, then $y = \lambda^i$ for some i.

Proof of Claim 5: There are at most r roots to the polynomial $u^r - 1$ in the field $Z(y)$. Since $\lambda^r = 1$, the elements $1, \lambda, \lambda^2, \cdots, \lambda^{r-1}$ are all distinct (because r is prime) and they are all roots of $u^r - 1$. So any $y^r = 1$ must be necessarily of the form $y = \lambda^i$ for some $0 \leq i \leq r-1$. □

Claim 6. $b^r \in Z$

Proof of Claim 6. Since $\lambda^r = 1$, and $\lambda \in Z$ we obtain

$$b^r = \lambda^r b^r = (\lambda b)^r = (a^{-1} b a)^r = a^{-1} b^r a$$

Note that here we used $\lambda b = a^{-1} b a$ which is equivalent to $ba = \lambda ab = \lambda ba$. We have shown that $ab^r = b^r a$. Since $ab \neq ba$ (because $ba = \lambda ab$ and $\lambda \neq 1$), by the result of **Claim 1** we get $b^r \in Z$. □

Let's recall from field theory that every finite multiplicative subgroup of a field is cyclic. In particular, the center Z is cyclic. Suppose γ is a generator for Z. We have $a^r \in Z$ and $b^r \in Z$ (the latter is from **Claim 6**). Hence, there exist integers j, k such that $a^r = \gamma^n$ and $b^r = \gamma^m$. If $r \mid n$, then $n = sr$ for some $s \in \mathbb{Z}$, and so $a^r = \gamma^{sr}$, which implies $(a/\gamma^s)^r = 1$; by **Claim 5**, it follows that $a/\gamma^s = \lambda^i \;\; \Rightarrow a = \gamma^s \lambda^i \in Z$ which contradicts the assumption that $a \notin Z$. We deduce that $r \nmid n$. Similarly, $r \nmid m$ (because $b \notin Z$). Let $a_1 = a^m$ and $b_1 = b^n$. From $ba = \lambda ab$, we have $bab^{-1} = \lambda a$. Since $\lambda \in Z$, we have

$$ba_1b^{-1} = ba^mb^{-1} = (bab^{-1})^m = (\lambda a)^m = \lambda^m a^m = \lambda^m a_1$$

So $ba_1b^{-1} = \lambda^m a_1 \;\; \Rightarrow \;\; ba_1 = \lambda^m a_1 b \;\; \Rightarrow \lambda^{-m} b = a_1 b a_1^{-1}$. Since $\lambda^{-m} \in Z$, we have

$$a_1 b_1 a_1^{-1} = a_1 b^n a_1^{-1} = (a_1 b a_1^{-1})^n = (\lambda^{-m} b)^n = \lambda^{-mn} b^n = \lambda^{-mn} b_1$$

Let $\mu = \lambda^{-mn} \in Z$. We have shown that $a_1 b_1 = \mu b_1 a_1$. Since $r \nmid m$, $r \nmid n$ and r is a prime number, it follows that $r \nmid mn$, meaning that $\lambda^{-mn} \neq 1 \;\; \Rightarrow \mu \neq 1$.

Note that $a_1^r = (a^m)^r = (a^r)^m = \gamma^{mn} = (b^r)^n = (b^n)^r = b_1^r$. So we have shown $a_1^r = b_1^r = gamma^{mn} \in Z$, and that $a_1 b_1 = \mu b_1 a_1$ for $mu \in Z$ with $\mu \neq 1$, but $\mu^r = 1$.

Claim 7. We have $(a_1^{-1} b_1)^r = \mu^{r(r-1)/2}$.

Proof of Claim 7. We claim that for any $t \in \mathbb{N}$,

$$(a_1^{-1} b_1)^t = \mu^{t(t-1)/2} a_1^{-t} b_1^t$$

We proceed by induction. When $t = 1$, the claim trivially holds, as both sides of the above equation reduce to $a_1^{-1} b_1$. Suppose that the statement holds for some $t \in \mathbb{N}$. We want to show that it holds for $t+1$ as well. It is instructive to write $t(t-1)/2 = 1 + 2 + \cdots + t - 1$. Then, using $a_1 b_1 = \mu b_1 a_1$, we have $b_1 a_1^{-1} = \mu a_1^{-1} b_1$, and so

$$\begin{aligned}
(a_1^{-1} b_1)^{t+1} &= a_1^{-1} (b_1 a_1^{-1})^t b_1 = a_1^{-1} (\mu a_1^{-1} b_1)^t b_1 \\
&= a_1^{-1} \mu^t (a_1^{-1} b_1)^t b_1 = a_1^{-1} \mu^t \mu^{t(t-1)/2} a_1^{-t} b_1^t b_1 \\
&= mu^{t+(t-1)+(t-2)+\cdots+2+1} a_1^{-(t+1)} b_1^{t+1} \\
&= \mu^{t(t+1)/2} a_1^{-(t+1)} b_1^{t+1}
\end{aligned}$$

proving the claim for $t+1$. By the principle of mathematical induction, the claim holds for all $t \in \mathbb{N}$. In particular, when $t = r$ in which case we obtain

$$(a_1^{-1} b_1)^r = \mu^{r(r-1)/2} a_1^{-r)} b_1^r = \mu^{r(r-1)/2}$$

So $\mu^{r(r-1)/2} = \mu^{r(r-1)/2}$. □

We divide the proof into cases, whether or not r is an odd prime. If r is an odd prime, then $\frac{r-1}{2}$ is an integer, and so $\mu^r = 1$ implies that $\mu^{r(r-1)/2} = 1$. By **Claim 7**, this means $(a_1^{-1} b_1)^r = 1$. Since $a_1^{-1} b_1$ satisfies $y^r = 1$, from **Claim 5** we obtain that $a_1^{-1} b_1 = \lambda^i$ for some i. So that $b_1 = a_1 \lambda^i = \lambda^i a_1$ (as $\lambda \in Z$). However, this implies $b_1 a_1 = a_1 b_1 = \mu b_1 a_1 \;\; \Rightarrow \;\; \mu = 1$, which contradicts $\mu \neq 1$. The proof of Wedderburn's theorem is finally complete when r is an odd prime.

Assume now $r = 2$. Then, let $\alpha = a_1^2 = b_1^2 \in Z$. Recall that $a_1 b_1 = \mu b_1 a_1$, where $\mu^2 = 1$, $\mu \neq 1$. Thus, $\mu = -1$, and $a_1 b_1 = -b_1 a_1 \neq b_1 a_1$. By **Lemma 1** in the beginning of the section, letting $F = Z$, we can find elements $\zeta, \eta \in Z$ such that $1 + \zeta^2 - \alpha \eta^2 = 0$. We will consider the expansion $(a_1 + \zeta b_1 + \eta a_1 b_1)^2$. Because of the condition $a_1 b_1 = -b_1 a_1$, the "cross" terms cancel, and we have are only left with "diagonal" terms as follows:

$$(a_1 + \zeta b_1 + \eta a_1 b_1)^2 = a_1^2 + \zeta^2 b_1^2 + \eta^2 (a_1 b_1)(a_1 b_1) = \alpha + \zeta^2 \alpha - \eta^2 \alpha^2 = \alpha(1 + \zeta^2 - \alpha \eta^2) = 0$$

Since D is a division ring, this means $a_1 + \zeta b_1 + \eta a_1 b_1 = 0$. But now

$$2a_1^2 = a_1(a_1 + \zeta b_1 + \eta a_1 b_1) + (a_1 + \zeta b_1 + \eta a_1 b_1) a_1 = 0$$

giving us $2a_1^2 = 0$, which forces $a_1 = 0$, contradiction. Wedderburn's theorem has been proved. □

References

[1] Adam, M. & Mutschler, B. J. "On Wedderburn's Theorem About Finite Division Algebras". Web link: http://www.uni-math.gwdg.de/mad/Wedderburn/Wedderburn.pdf

[2] Herstein, I. N. "Wedderburn's theorem and a theorem of Jacobson." *Amer. Math. Monthly* **68** 1961 249 - 251.

[3] Kaczynski, T. J. "Another Proof of Wedderburn's Theorem." *Amer. Math. Monthly* **71** 1964 pp. 652-653

[4] Rotman, Joseph J. *Advanced modern algebra.* Second edition. Graduate Studies in Mathematics, 114. American Mathematical Society, Providence, RI, 2010.

[5] Zassenhaus, Hans J. "A group-theoretic proof of a theorem of Maclagan-Wedderburn." *Proc. Glasgow Math. Assoc.* **1** (1952) 53 - 63.

[6] Wedderburn, J. H. M. "A theorem on finite algebras." *Trans. Amer. Math. Soc.* **6** (1905), no. 3, 349352.