Math. Proc. Camb. Phil. Soc. (1978), **84**, 455
Printed in Great Britain

On length functions and normal forms in groups

By BERNARD HURLEY

Department of Mathematics, The City University, St John's Street, London EC1V 4PB

(*Received* 19 *August* 1977, *revised* 14 *November* 1977)

1. *Introduction.* Abstract length functions were first considered axiomatically in Lyndon(4). Length functions satisfying a less restrictive set of axioms were studied by Chiswell(1). Groups with a normal form structure (NFS-groups) were studied in Hurley(3) using a slightly different axiomatization to that used in this paper. Length functions naturally arise from normal form structures. In this paper we show that the length functions arising from normal form structures can be characterized axiomatically by a set of axioms that consists of those of (1) together with one further axiom.

2. *NFS-groups.* In this section we consider groups that are equipped with a normal form structure satisfying certain axioms. Before stating these axioms we need some definitions. For most of this section, G is a fixed group, X is a fixed subset of G, satisfying $1 \notin X$, and B is a fixed subgroup of G called the *base group*.

Definitions. A (G, B, X)-*sequence* $\tau = (\tau_i)$ of length $n \geqslant 0$ is a sequence of $2n+1$ (not n) elements of G such that $\tau_{2j+1} \in B$ for $j = 0, \ldots, n$ and $\tau_{2j} \in X \cup X^{-1}$ for $j = 1, \ldots, n$. The set of all such sequences shall be denoted by $\mathbf{S}(G, B, X)$ or just $\mathbf{S}$. If $\sigma \in \mathbf{S}$, then $|\sigma|$ shall denote its length and $\overline{\sigma}$ shall denote the product of its elements taken in their natural order.

If $\tau = (\tau_i) \in \mathbf{S}$ and for some $1 \leqslant j \leqslant n-1$, $\tau_{2j}\tau_{2j+1}\tau_{2j+2} \in B$, then the sequence σ of length $n-2$ defined by:

$$\sigma_i = \begin{cases} \tau_i & \text{if} \quad 1 \leqslant i \leqslant 2j-2, \\ \tau_{2j-1}\tau_{2j}\tau_{2j+1}\tau_{2j+2}\tau_{2j+3} & \text{if} \quad i = 2j-1, \\ \tau_{i+4} & \text{if} \quad 2j \leqslant i \leqslant 2n-4, \end{cases}$$

is clearly in $\mathbf{S}$ and satisfies $\overline{\sigma} = \overline{\tau}$. We say that σ is obtained from τ by means of a *pinch* between τ_{2j} and τ_{2j+2}.

If $\tau = (\tau_i) \in \mathbf{S}$ and for some $1 \leqslant j \leqslant n$, $\tau_{2j}\tau_{2j+1}\tau_{2j+2} \in B(X \cup X^{-1})B$, then

$$\tau_{2j}\tau_{2j+1}\tau_{2j+2} = bgc$$

for some $b, c \in B$ and for some $g \in X \cup X^{-1}$ and the sequence μ of length $n-1$ defined by:

$$\mu_i = \begin{cases} \tau_i & \text{if} \quad 1 \leqslant i \leqslant 2j-2, \\ \tau_{2j-1}b & \text{if} \quad i = 2j-1, \\ g & \text{if} \quad i = 2j, \\ c\tau_{2j+3} & \text{if} \quad i = 2j+1, \\ \tau_{i+2} & \text{if} \quad 2j+2 \leqslant i \leqslant 2n-2, \end{cases}$$

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

is clearly in **S** and satisfies $\bar{\mu} = \bar{\tau}$. We say that μ is obtained from τ by means of an *amalgamation* between τ_{2j} and τ_{2j+2}.

A (G, B, X)-sequence τ is *reduced* if no pinch or amalgamation can be performed between any two elements of τ. The set of all such sequences shall be denoted by $\mathbf{RS}(G, B, X)$ or just **RS**. Clearly if $\sigma \in \mathbf{S}$ then there exists $\tau \in \mathbf{RS}$ such that $\bar{\sigma} = \bar{\tau}$.

We say that $\sigma, \tau \in \mathbf{S}$ are *parallel*, written $\sigma \| \tau$, if, either $|\sigma| = |\tau| = 0$ and $\bar{\sigma} = \bar{\tau}$, or $|\sigma| = |\tau| = n > 0$ and there exist $h_i \in B$ for $i = 1, 2, \ldots, 2n$ such that $\tau_1 = \sigma_1 h_1$,

$$\tau_i = h_{i-1}^{-1} \sigma_i h_i \quad \text{for} \quad i = 2, \ldots, 2n, \quad \text{and} \quad \tau_{2n+1} = h_{2n}^{-1} \sigma_{2n+1}.$$

It is easy to see that if $\sigma \| \tau$ then $\bar{\sigma} = \bar{\tau}$.

We introduce the axioms:

N1. $G = gp(X, B)$.

N2. If $\sigma, \tau \in \mathbf{RS}(G, B, X)$ and $\bar{\sigma} = \bar{\tau}$, then $\sigma \| \tau$.

We call a triple (G, B, X) satisfying **N1** and **N2** an *NFS-group* (NFS = normal form structure). Alternatively we say (G, B, X) is an *NF-structure* on G.

These axioms are easy to motivate. **N1** says that every element $g \in G$ can be expressed as a product of elements of $X \cup X^{-1} \cup B$. It easily follows that there exists $\sigma \in RS$ such that $g = \bar{\sigma}$; we say that this sequence *represents* g. Thus **N1** says that every element has a normal form. **N2** says that this normal form is unique up to parallelism.

At the suggestion of the referee, we shall now illustrate our axioms with some examples of NFS-groups. We leave it to the reader to check that the structures described really are NFS-groups; only example 4 presents any real difficulty. The reader might also like to check that in examples 2 and 3 only pinching is possible.

Example 1. G is any group, B is any subgroup of G and $X = G - B$.

Example 2. G is a free group freely generated by $X = \{t_1, t_2, \ldots, t_r\}$, and $B = 1$.

Example 3. $G = \langle K, t; t^{-1}Lt = M \rangle$, an H.N.N. extension, $B = K$ and $X = \{t\}$.

Example 4. $G = \langle H * K; U = V \rangle$, a free product with amalgamation, $B = U$ and $X = (H - U) \cup (K - V)$.

Chiswell (1) considers integer-valued length functions $(G \to \mathbb{Z}, g \mapsto |g|$ on a group $G)$ satisfying the following three axioms:

A1′. $|1| = 0$.

A2. $|x| = |x^{-1}|$.

A4. $d(x, y) < d(x, z)$ implies $d(y, z) = d(x, y)$.

$d(x, y)$ is defined to be $\frac{1}{2}(|x| + |y| - |xy^{-1}|)$. The numbering of the axioms is that of Lyndon (4). We introduce the new axiom:

N1*. $G = gp\{x \in G : |x| \leqslant 1\}$.

We define a length function to be *well-based* if it satisfies these four axioms.

Later in this section we shall show how a length function $G \to \mathbb{Z}$ can always be associated with an NF-structure (G, B, X) on G. We now state our main result, which we shall prove in Section 3.

MAIN THEOREM. *A function $G \to \mathbb{Z}$, $g \mapsto |g|$ on a group is a well-based length function if and only if it is the associated length function of some NF-structure on G.*

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

In many cases (e.g. Example 4 above), axiom **N2** is difficult to check directly. In an attempt to overcome this difficulty, we introduce some equivalent axiomatizations of NFS-groups. Defining $X_1 = B(X \cup X^{-1})B$, we introduce the axioms:

N3. If $\sigma \in \mathbf{RS}$ and $\overline{\sigma} = 1$, then $\sigma = (1)$.

N3*. If $\sigma \in \mathbf{S}$ and $\overline{\sigma} = 1$, then either $\sigma = (1)$ or there exists $\tau \in \mathbf{S} - \mathbf{RS}$ such that $\sigma \| \tau$.

N4. If $g, h, k, gh \in X_1$ but $hk \notin B \cup X_1$, then $ghk \notin B \cup X_1$.

N4*. If $g, h, k, hk \in X_1$ but $gh \notin B \cup X_1$, then $ghk \notin B \cup X_1$.

Before considering questions of the interdependence of our axioms, we prove some technical results.

LEMMA 2·1. (i) *Suppose* (G, B, X) *satisfies* **N4** *and* **N4*** *and* $\tau \in \mathbf{S}$. *If no pinches and precisely one amalgamation can be performed in* τ, *then the sequence* σ *obtained from* τ *by performing this amalgamation is reduced.*

(ii) *For any* (G, B, X), *if* $\sigma \| \tau$ *then* $\sigma \in \mathbf{RS}$ *if and only if* $\tau \in \mathbf{RS}$.

Proof. (i) Suppose that the amalgamation can be performed between τ_{2i} and τ_{2i+2}. We define $g = \tau_{2i-1}\tau_{2i}\tau_{2i+1}$, $h = \tau_{2i+2}\tau_{2i+3}$ and $k = \tau_{2i+4}$ so that $g, h, k \in X_1$. Since σ is obtained by performing this amalgamation, $gh = \sigma_{2i-1}\sigma_{2i}\sigma_{2i+1} \in X_1$. Since τ cannot be pinched or amalgamated except for the amalgamation between τ_{2i} and τ_{2i+2}, we see that $hk \notin B \cup X_1$. From **N4** we see that $ghk = \sigma_{2i-1}\sigma_{2i}\sigma_{2i+1}\sigma_{2i+2} \notin B \cup X_1$, and, since $\sigma_{2i-1} \in B$, we see $\sigma_{2i}\sigma_{2i+1}\sigma_{2i+2} \notin B \cup X_1$. Thus σ cannot be pinched or amalgamated between σ_{2i} and σ_{2i+2}. Similarly, using **N4***, it cannot be pinched or amalgamated between σ_{2i-2} and σ_{2i}. It is clear from our assumptions on τ, that σ cannot be pinched or amalgamated anywhere else. Hence σ is reduced.

(ii) This part is immediate from the definitions. ∎

Definition. If $\sigma, \tau \in \mathbf{S}$ where $|\sigma| = n$ and $|\tau| = m$ then we define sequences σ^{-1} and $\sigma * \tau$ by

$$(\sigma^{-1})_i = (\sigma_{2n+2-i})^{-1} \quad \text{for} \quad 1 \leqslant i \leqslant 2n+1,$$

and

$$(\sigma * \tau)_i = \begin{cases} \sigma_i & \text{if} \quad 1 \leqslant i \leqslant 2n, \\ \sigma_{2n+1}\tau_1 & \text{if} \quad i = 2n+1, \\ \tau_{i-2n} & \text{if} \quad 2n+2 \leqslant i \leqslant 2(m+n)+1. \end{cases}$$

We easily see that $|\sigma^{-1}| = |\sigma|$, $\overline{\sigma^{-1}} = (\overline{\sigma})^{-1}$, $(\sigma^{-1})^{-1} = \sigma$, $|\sigma * \tau| = |\sigma| + |\tau|$,

$$\overline{\sigma * \tau} = \overline{\sigma} * \overline{\tau} \quad \text{and} \quad (\sigma * \tau) * \mu = \sigma * (\tau * \mu).$$

From the last of these equations, the notation $\sigma * \tau * \mu$ is unambiguous.

LEMMA 2·2. *Axiom* **N3** *is equivalent to* **N3***. *Axiom* **N4** *is equivalent to* **N4***. *Axiom* **N2** *is equivalent to the pair* **N3** *and* **N4**. *Hence* (G, B, X) *is an NFS-group if and only if it satisfies axioms* **N1**, **N3** *and* **N4**.

Proof. The equivalence of **N3** and **N3*** follows immediately from Lemma 2·1 (ii). The equivalence of **N4** and **N4*** follows from the fact that X_1 and $B \cup X_1$ are closed under taking inverses.

We now turn to the equivalence of **N2** with the pair **N3** and **N4**. First assume **N2** is

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

satisfied. Clearly **N3** is also satisfied. If $g, h, k, gh \in X_1$ but $kh \notin B \cup X_1$, then there are sequences $\mu, \nu, \sigma, \tau \in \mathbf{RS}$ of length 1 such that $g = \overline{\mu}, h = \overline{\nu}$, $k = \overline{\sigma}$ and $gh = \overline{\tau}$ and $\alpha = (\nu_1, \nu_2, (\nu_3 \sigma_1), \sigma_2, \sigma_3) \in \mathbf{RS}$. If $ghk = b \in B$, then $g^{-1}hk \in X_1$, which is contradictory. If $ghk \in X_1$, then there exists $\gamma \in \mathbf{RS}$ of length 1 such that $ghk = \overline{\gamma}$. Putting

$$\delta = (\mu_3^{-1}, \mu_2^{-1}, (\mu_1^{-1}\gamma_1), \gamma_2, \gamma_3) \in \mathbf{S},$$

we see that $\overline{\delta} = g^{-1}ghk = hk = \overline{\alpha}$. Now $|\alpha| = |\delta|$ and so we see $\delta \in \mathbf{RS}$ and, from **N2**, that $\alpha \| \delta$. Hence $\nu_1 = \mu_3^{-1} h_1, \nu_2 = h_1^{-1} \mu_2^{-1} h_2$ for some $h_1, h_2 \in B$. Thus

$$gh = \mu_1 \mu_2 \mu_3 \nu_1 \nu_2 \nu_3 = \mu_1 h_2 \nu_3 \in B,$$

which contradicts $gh \in X_1$. Thus **N4** holds.

Now suppose that **N3** and **N4** hold and suppose $\sigma, \tau \in \mathbf{RS}$ satisfy $\overline{\sigma} = \overline{\tau}$. We let $n = |\sigma|$ and $m = |\tau|$ and we show that **N2** holds by induction on $\min(n, m)$. If

$$\min(n, m) = 0,$$

suppose $m = 0$ (the case $n = 0$ is similar), then defining $\mu = \sigma * (\tau^{-1})$ we see $\mu = 1$ and $\mu \in \mathbf{RS}$ and so, by **N3**, $\mu = (1)$. Hence $\sigma = (\sigma_1) = \tau = (\tau_1)$ and so $\sigma \| \tau$.

If $\min(n, m) > 0$, then clearly $n \geqslant 1$ and $m \geqslant 1$. Defining $\mu = \sigma * (\tau^{-1})$, we see that $\overline{\mu} = 1$ and so by **N3**, $\mu \notin \mathbf{RS}$. This is only possible if a pinch or an amalgamation can be performed between $\mu_{2n} = \sigma_{2n}$ and $\mu_{2n+2} = \tau_{2m}^{-1}$. If an amalgamation can be performed at this point, then, by Lemma 2·1 (i), the sequence ζ obtained from μ by performing this amalgamation is reduced. But then, by **N3**, $\zeta = (1)$. Thus $0 = |\zeta| = |\mu| - 1 = m + n - 1$ which contradicts $m \geqslant 1$ and $n \geqslant 1$.

Now suppose a pinch can be performed at this point, then let ν be the sequence of length $m + n - 2$ obtained by pinching at this point. If we now define $\sigma', \tau' \in \mathbf{S}$ of lengths $n - 1$ and $m - 1$ respectively by:

$$\sigma_i' = \begin{cases} \sigma_i & \text{if} \quad 1 \leqslant i \leqslant 2n-2 \\ \nu_{2n-1} & \text{if} \quad i = 2n-1, \end{cases}$$

and

$$\tau_i' = \begin{cases} \tau_i & \text{if} \quad 1 \leqslant i \leqslant 2m-2 \\ 1 & \text{if} \quad i = 2m-1, \end{cases}$$

then clearly $\sigma', \tau' \in \mathbf{RS}$ and $\overline{\sigma}' = \overline{\tau}'$. By our induction hypothesis $\sigma' \| \tau'$. It easily follows that $\sigma \| \tau$. Thus **N2** holds. ∎

Definition. If (G, B, X) is an NFS-group, then we define the *associated length function* on G by putting $|g| = |\sigma|$ for all $g \in G$ where $\sigma \in \mathbf{RS}$ represents g.

Remark. It is clear from **N1** and **N2** that this function is well defined and has non-negative integer values. It is also clear that $B = \{x \in G : |x| = 0\}$, $X_1 = \{x \in G : |x| = 1\}$ and $|x| = |x^{-1}|$ for all $x \in G$.

Many important concepts can be defined entirely in terms of the length function. For instance we define the *amount of agreement on the right function* by

$$d(x, y) = \tfrac{1}{2}(|x| + |y| - |xy^{-1}|).$$

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

This is clearly either an integer or half an integer. We say that *no cancellation occurs* in the product xy if $d(x, y^{-1}) = 0$ and *almost no cancellation* occurs in the product xy if $d(x, y^{-1}) \leqslant \frac{1}{2}$. We say that two NF-structures on G are *isometric* if they give rise to the same length function. Clearly this defines an equivalence relation on the NF-structures on G and we call the equivalence classes *isometry classes*.

Definitions. An NFS-group (G, B, X) is *full* if it satisfies the axiom:

$$\mathbf{F}.\quad X = B(X \cup X^{-1})B \quad (\text{i.e. } X = X_1).$$

An NFS group (G, B, X) is *sparse* if it satisfies the axiom:

$$\mathbf{SP}.\ \text{If}\quad xay^{\epsilon}b = 1 \quad \text{where} \quad a, b \in B; x, y \in X \quad \text{and} \quad \epsilon = \pm 1, \quad \text{then} \quad x = y.$$

If G is a group, and A, B, X, Y are subsets of G, we write $(G, B, X) \leqslant (G, A, Y)$ and say that (G, B, X) is *contained in* (G, A, Y), if $B \subseteq A$ and $X \subseteq Y$. This clearly induces a partial order on the NF-structures on G.

LEMMA 2·3. (i) *(G, B, X) and (G, A, Y) are isometric NF-structures on a group G if and only if $B = A$ and $X_1 = Y_1$.*

(ii) *If $(G, A, X) \leqslant (G, B, Y) \leqslant (G, C, Z)$ and (G, A, X) and (G, C, Z) are isometric NF-structures, then (G, B, Y) is also an NF-structure and is isometric to the other two.*

(iii) *Every isometry class of NF-structures on a group G contains a unique maximal element which is full. Every full NF-structure is the maximal element of the isometry class to which it belongs.*

(iv) *The minimal NF-structures of an isometry class are sparse and every sparse NF-structure is a minimal element of the isometry class to which it belongs. Every NF-structure contains a sparse NF-structure isometric to it.*

Proof. (i) By the remark following the definition of the length function, it follows that if (G, B, X) and (G, A, Y) are isometric NF-structures then $B = A$ and $X_1 = Y_1$.

Now suppose $A = B$ and $X_1 = Y_1$. Let $g \in G$ and let $\sigma \in \mathbf{RS}(G, B, X)$ represent g. If $|\sigma| = n$, then $\sigma_{2i} \subseteq X \cup X^{-1} \subseteq Y_1$ for $i = 1, \ldots, n$. Hence there exist $\tau_{2i} \in Y \cup Y^{-1}$ and $b_i, c_i \in B$ such that $\sigma_{2i} = b_i \tau_{2i} c_i$ for $i = 1, \ldots, n$. If we define $\tau_{2i+1} = c_i \sigma_{2i+1} b_{i+1}$ for $i = 0, \ldots, n$, where $c_0 = b_{n+1} = 1$ then we see that $\tau \in \mathbf{S}(G, B, Y)$. We see easily that τ is reduced and represents g. Since $|\tau| = n$, the two NF-structures are isometric.

(ii) If $(G, A, X) \leqslant (G, B, Y) \leqslant (G, C, Z)$ and (G, A, X) is isometric to (G, C, Z) then $A \subseteq B \subseteq C = A$ and so $A = B = C$. Moreover

$$X_1 = B(X \cup X^{-1})B \subseteq B(Y \cup Y^{-1})B = Y_1 \subseteq B(Z \cup Z^{-1})B = Z_1 = X_1$$

and so $X_1 = Y_1 = Z_1$. We easily see that (G, B, Y) is an NF-structure. By (i), all three NF-structures are isometric.

(iii) It is clearly sufficient to prove that given any NF-structure (G, B, X), (G, B, X_1) is a full NF-structure. This is a routine check of axioms **N1** and **N2**.

(iv) Suppose $(G, A, Y) \lesneqq (G, B, X)$ and suppose that (G, B, X) is a sparse NF-structure on G. To show that (G, B, X) is minimal in its isometry class we must show that these structures are not isometric. This is obviously so if $A \neq B$ so suppose

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

$A = B$. Then $X - Y$ must be non-empty. Pick $x \in X - Y$ and let $\sigma \in \mathbf{RS}(G, A, Y)$ represent x. Clearly $|\sigma| \geqslant 1$. If $|\sigma| = 1$, then there exists $y \in Y$, $a, b \in B$ and $\epsilon = \pm 1$ such that $xay^{\epsilon}b = 1$ in G contradicting (G, B, X) being sparse. Thus $|\sigma| > 1$ and so the two structures are not isometric. Thus (G, B, X) is a minimal element of its isometry class.

If (G, B, X) is any NF-structure on G, define an equivalence relation $\sim$ on X by saying $x \sim y$ if and only if there exist $a, b \in B$ and $\epsilon = \pm 1$ such that $xby^{\epsilon}c = 1$ in G. If Y is a set of representatives for the equivalence classes on X, then (G, B, Y) is a sparse NF-structure isometric to and contained in (G, B, X). |

In the introduction to Lyndon (4), it is pointed out that while, intuitively, the function $d(x, y)$ measures the amount of agreement on the right between x and y, it does not follow that x and y contain their greatest right common divisor. The next lemma shows that in our situation they do.

LEMMA 2·4. *Let (G, B, X) be an NFS-group and let $x, y \in G$. Suppose that*

$$d(x, y) = r + \eta,$$

where $r \geqslant 0$ is an integer and $\eta = 0$ or $\frac{1}{2}$. Then there exist $x_1, y_1, t \in G$ such that $|t| = r$, $x = x_1 t$ and $y = y_1 t$ with no cancellation occurring in these products and

$$|x_1 y_1^{-1}| = |x_1| + |y_1| - 2\eta.$$

Proof. By induction on r. If $r = 0$ then putting $x_1 = x$, $y_1 = y$ and $t = 1$ the result follows.

Now suppose $r > 0$ and suppose inductively that the lemma holds for all pairs $z, w \in G$ with $d(z, w) < r$. Let $\sigma, \tau \in \mathbf{RS}$ represent x and y respectively, suppose $|x| = n$, $|y| = m$ and define $\mu = \sigma * (\tau^{-1})$. Since $r > 0$, it follows, from Lemma 2·1 (i), that

$$\mu_{2n}\mu_{2n+1}\mu_{2n+2} = \sigma_{2n}\sigma_{2n+1}\tau_{2m+1}^{-1}\tau_{2m}^{-1} \in B.$$

Putting $u = \tau_{2m}\tau_{2m+1}$ we see that:

$$\zeta = (\sigma_1, \sigma_2, \ldots, \sigma_{2n-2}, (\sigma_{2n-1}\sigma_{2n}\sigma_{2n+1}\tau_{2m+1}^{-1}\tau_{2m}^{-1}))$$

and

$$\xi = (\tau_1, \tau_2, \ldots, \tau_{2m-2}, \tau_{2m-1})$$

are both reduced and represent xu^{-1} and yu^{-1} respectively. Thus $|xu^{-1}| = |x| - 1$ and $|yu^{-1}| = |y| - 1$ and so

$$\begin{aligned} d(xu^{-1}, yu^{-1}) &= \tfrac{1}{2}\{|xu^{-1}| + |yu^{-1}| - |(xu^{-1})(yu^{-1})^{-1}|\} \\ &= \tfrac{1}{2}\{|x| - 1 + |y| - 1 + |xy^{-1}|\} \\ &= r - 1 + \eta. \end{aligned}$$

By induction, there exist $x_1, y_1, v \in G$ such that $|v| = r - 1$, $xu^{-1} = x_1 v$, $yu^{-1} = y_1 v$, with no cancellation occurring on the right-hand sides of these equations, and

$$|x_1 y^{-1}| = |x_1| + |y_1| - 2\eta.$$

Put $t = vu$. Let $\rho, \chi \in \mathbf{RS}$ represent y_1 and v respectively. Since no cancellation occurs in the product $y_1 v$, it follows that $\rho * \chi$ is reduced and represents $y_1 v = yu^{-1}$. Thus, by

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

N2, $\xi \| \rho * \chi$. Thus there exists $k \in B$ such that $\chi_{2r-1} = k\tau_{2m-1}$. (N.B. If $r = 1$, put $k = 1$.) Since τ is reduced, we see that $\chi_{2r-1}\tau_{2m}\tau_{2m+1} = k\tau_{2m-1}\tau_{2m}\tau_{2m+1} \notin B \cup X_1$. Hence

$$\lambda = (\chi_1, \chi_2, \cdots, \chi_{2r-2}, \chi_{2r-1}, \tau_{2m}, \tau_{2m+1})$$

is reduced. Clearly $|\lambda| = r$ and $\bar{\lambda} = t$ and so $|t| = r$. Moreover with $z = x_1$ or y_1 we see

$$\begin{aligned} d(z, t^{-1}) &= \tfrac{1}{2}\{|z| + |t^{-1}| - |zt|\} \\ &= \tfrac{1}{2}\{|z| + |v^{-1}| + 1 - |zv| - 1\} \\ &= d(z, v^{-1}) = 0. \end{aligned}$$

Thus no cancellation occurs in the products $x = x_1 t$ and $y = y_1 t$ and the elements $x_1, y_1, t \in G$ are the required elements of G. ∎

The next lemma follows easily from our normal form axioms and its proof will be omitted.

LEMMA 2·5. *Let (G, B, X) be an NFS-group and suppose $x, y, z, u \in G$ satisfy $|x| \leqslant |z|$ and $xy = zu$ with no cancellation occurring in either product. Then there exists $t \in G$ such that $z = xt$ and $y = tu$, no cancellation occurring in either product.*

3. *The associated length function.* The bulk of this section will be devoted to proving the main theorem.

LEMMA 3·1. *The associated length function of an NFS-group is well-based.*

Proof. Axioms **A1′**, **A2** and **N1*** are clearly satisfied and so it is sufficient to show that axiom **A4** is satisfied. Suppose then that (G, B, X) is our NFS-group and $|-|$ is its associated length function. Suppose we have $x, y, z \in G$ such that $d(x, y) < d(x, z)$. Putting

$$r = [d(x, y)] \quad \text{and} \quad s = [d(x, z)],$$

we see that $r \leqslant s$ and, by Lemma 2·4, there exists $x_1, x_2, t_1, t_2, y_1, z_2 \in G$ such that $|t_1| = r$, $|t_2| = s$, $x = x_1 t_1 = x_2 t_2$, $y = y_1 t_1$ and $z = z_2 t_2$, with no cancellation occurring in any of these products. Moreover $|x_1 y_1^{-1}| = |x_1| + |y_1| - 2\eta_1$ and $|x_2 z_2^{-1}| = |x_2| + |z_2| - 2\eta_2$, where $\eta_1, \eta_2 = 0$ or $\frac{1}{2}$. There are now two cases.

Case 1. $r < s$. Applying Lemma 2·5 to the equation $x_1 t_1 = x_2 t_2$, we see that there exists $u \in G$ such that $|u| \geqslant 1$, $x_1 = x_2 u$ and $t_2 = u t_1$ with no cancellation occurring in either product.

Representing the elements of G by sequences in **RS**, and using Lemma 2·1 (i), we easily see that $|y_1 u^{-1} z_2^{-1}| = |y_1| + |u| + |z_2| - 2\eta_1$. It follows that

$$d(y, z) = |t_1| + \eta_2 = d(x, y).$$

Case 2. $r = s$. This can only occur if $d(x, y) = r$ and $d(x, z) = r + \frac{1}{2}$. Clearly $|x_1| \geqslant 1$ and $|z_2| \geqslant 1$ and we may suppose $t_1 = t_2$ and $x_1 = x_2$. If $|y_1| = 0$ then obviously

$$d(y, z) = d(x, y);$$

so suppose $|y_1| \geqslant 1$. Let $\sigma, \tau, \mu \in \mathbf{RS}$ represent x_1, y_1 and z_2 respectively and suppose $|x_1| = a$, $|y_1| = b$ and $|z_2| = c$.

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

Now precisely one amalgamation can be performed in $\sigma * (\mu^{-1})$, so that for some $h, k \in B$ and some $u \in X \cup X^{-1}$, we have:

$$\sigma_{2a-1}\sigma_{2a}\sigma_{2a+1}\mu_{2c+1}^{-1}\mu_{2c}^{-1}\mu_{2c-1}^{-1} = huk. \tag{1}$$

We now define

$$\lambda = (\tau_{2b-1}, \tau_{2b}, (\tau_{2b+1}\sigma_{2a+1}^{-1}), \sigma_{2a}^{-1}, (\sigma_{2a-1}^{-1}h), u, k).$$

Now $\sigma * (\tau^{-1}) \in \mathbf{RS}$ and it follows that no pinch or amalgamation can be performed between τ_{2b} and σ_{2a}^{-1} in λ. However, from equation (1), we see that an amalgamation can be performed between σ_{2a}^{-1} and u. The sequence obtained by performing this amalgamation is $(\tau_{2b-1}, \tau_{2b}(\tau_{2b+1}\mu_{2c+1}^{-1}), \mu_{2c}^{-1}, (\mu_{2c-1}^{-1}))$ and from Lemma 2·1(i) it is reduced. Hence $\tau * (\mu^{-1})$ must be reduced. So no cancellation occurs in the product $y_1 z_2^{-1}$. Hence $d(y, z) = d(x, y)$, as required. |

To complete the proof of the main theorem we need to prove a converse to Lemma 3·1. To this end we shall, for the time being, assume we are dealing with a group G equipped with a length function $g \mapsto |g|$ satisfying **A1′**, **A2** and **A4**. We shall impose axiom **N1*** later. We write $B = \{x \in G : |x| = 0\}$ and $X = \{x \in G : |x| = 1\}$. Taking propositions 5 and 6 of Hoare(2) together with axiom **A1′** we immediately see:

LEMMA 3·2. *For all $x, y \in G$, $d(x, y) \geqslant 0$, $|x| \geqslant 0$ and B is a subgroup of G.*

LEMMA 3·3. *For all $x, y \in G$, $\||x| - |y|\| \leqslant |xy| \leqslant |x| + |y|$.*

Proof. The right-hand inequality is merely $d(x, y^{-1}) \geqslant 0$ rewritten slightly with the aid of **A2.** A standard manipulation now yields the left-hand inequality.

COROLLARY. *The actions of B on G by left and right multiplication are length-preserving. That is if $x \in G$ and $b \in B$ then $|xb| = |bx| = |x|$.*

Proof. $|x| = \||x| - |b|\| \leqslant |xb| \leqslant |x| + |b| = |x|$. Hence $|x| = |xb|$ and similarly $|x| = |bx|$.

Part (i) of the next lemma tells us that if, in a product uv, u cancels entirely into v, then u cancels with some initial segment of v. Part (ii) tells us how words of length one can amalgamate with longer words. Before stating and proving the lemma we note that axiom **A4** is equivalent to:

A4*. If $d(x, y) \geqslant m$ and $d(y, z) \geqslant m$ then $d(x, z) \geqslant m$.

LEMMA 3·4. (i) *If $x, y, z \in G$ satisfy $|x| = |y|$, $|yz| = |y| + |z|$ and $|xyz| = |z|$, then $|xy| = 0$.*

(ii) *If $x, y, z \in G$ satisfy $|x| = |y| = 1$ and $|yz| = |xyz| = 1 + |z|$ then $|xy| = 1$.*

Proof. (i) Suppose $|x| = p$ and $|z| = q$, then: $d(x, z^{-1}y^{-1}) = d(z^{-1}y^{-1}, y^{-1}) = p$. By **A4***, we have $d(x, y^{-1}) \geqslant p$ and hence $|xy| \leqslant 0$. Thus, from Lemma 3·2, we see $|xy| = 0$.

(ii) In this case: $d(x, z^{-1}y^{-1}) = \frac{1}{2} < d(z^{-1}y^{-1}, y^{-1}) = 1$. So, by **A4**, we have

$$d(x, y^{-1}) = \tfrac{1}{2}$$

and hence $|xy| = 1$.

We now come to the converse of Lemma 3·1.

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

LEMMA 3·5. *Let G be a group with a well-based integer-valued length function $G \to \mathbb{Z}$, $g \mapsto |g|$. If $B = \{g \in G: |g| = 0\}$ and $X = \{g \in G: |g| = 1\}$ then (G, B, X) is a full NFS-group with associated length function $g \mapsto |g|$.*

Proof. It is clear that if (G, B, X) is an NFS-group then it is full. **N1** is clearly satisfied so, by Lemma 2·2, we need to check axioms **N3** and **N4**. Since **N1** holds, each $x \in G$ can be represented by some $\sigma \in \mathbf{RS}(G, B, X)$. We now show:

(*) If $\sigma \in \mathbf{RS}(G, B, X)$ represents $x \in G$, then $|\sigma| = |x|$.

In (*), $|x|$ means the length of x under the given length function while $|\sigma|$ means the length of σ as defined in Section 2. We can write: $x = h_0 y_1 h_1 y_2 \dots h_{|\sigma|-1} y_{|\sigma|} h_{|\sigma|}$, for some $h_0, \dots, h_{|\sigma|} \in B$ and $y_1, \dots, y_{|\sigma|} \in X$. So, by Lemma 3·3, we see:

$$|x| \leqslant \sum_{i=0}^{|\sigma|} |h_i| + \sum_{i=1}^{|\sigma|} |y_i| = |\sigma|.$$

If equality does not hold for all pairs $x \in G$ and $\sigma \in \mathbf{RS}$ such that $x = \overline{\sigma}$, then we may suppose that we are dealing with a pair (x, σ) such that $x \neq |\sigma|$ and $|\sigma|$ is minimal with respect to this property. Let $y = h_0 y_1 \dots h_{|\sigma|-1}$. By the minimality of $|\sigma|$, we see that: $|y| = |\sigma| - 1$. By the corollary to Lemma 3·3, we see: $|y_{|\sigma|} h_{|\sigma|}| = |y_{|\sigma|}| = 1$, and we see: $\|\sigma| - 2| = \|y| - 1| \leqslant |x| < |\sigma|$. Hence $|\sigma| > 1$ and $|x| = |\sigma| - 1$ or $|\sigma| - 2$. From an obvious variant of Lemma 3·4, we see:

$$|y_{|\sigma|-1} h_{|\sigma|-1} y_{|\sigma|} h_{|\sigma|}| = \begin{cases} 0 & \text{if} \quad |x| = |\sigma| - 2, \\ 1 & \text{if} \quad |x| = |\sigma| - 1. \end{cases}$$

In either case this contradicts $\sigma \in \mathbf{RS}$. Hence (*) holds.

It follows from (*) that **N3** holds and that if (G, B, X) is an NF-structure on G, then its associated length function is $g \mapsto |g|$.

To complete the proof we must show that **N4** holds. Suppose $g, h, k, gh \in X$ and $hk \notin B \cup X$. Then $|g| = |h| = |k| = |gh| = 1$ and, by (*), $|hk| = 2$. Hence

$$1 = \|hk| - |g\| \leqslant |ghk|.$$

If $|ghk| = 1$, then, by Lemma 3·4(i), we see $|gh| = 0$ – a contradiction. Thus we see $|ghk| > 1$, and hence $|ghk| \notin B \cup X$ as required. ∎

The main theorem now follows immediately from Lemmas 3·1 and 3·5. We finish off this paper by considering the effects of imposing more axioms on an NFS-group. Consider Lyndon's axioms:

A1. $|x| = 0$ if and only if $x = 1$.

A5. $d(x, y) + d(x^{-1}, y^{-1}) > |x| = |y|$ implies that $x = y$.

LEMMA 3·6. *Let (G, B, X) be an NFS-group and let $|-|$ be its associated length function. Then*:

(i) *$|-|$ satisfies* **A1** *if and only if $B = 1$.*

(ii) *If $B = 1$ then $|-|$ satisfies* **A5**.

Proof. (i) is clear.

(ii) If $B = 1$, then every sequence in **RS** has the form: $(1, u_1, 1, \dots, u_n, 1)$, where

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press

$u_i \in X \cup X^{-1}$ for $i = 1, \ldots, n$. We call $u_1, \ldots, u_n$ the syllables of the sequence. Suppose $x, y \in G$ are represented by $\sigma, \tau \in \mathbf{RS}$ respectively and satisfy

$$d(x, y) + d(x^{-1}, y^{-1}) > |x| = |y|.$$

Then we see, from Lemma 2·4, that the number of syllables in which σ and τ agree on the right plus the number in which they agree on the left adds up to a number not smaller than their common length. Hence $\sigma = \tau$ and hence $x = y$.

Example. If we let $G = \langle a, b \rangle$, the free group on a and b, $B = \langle b \rangle$ and $X = \{a\}$ then it is easy to see that (G, B, X) is an NFS-group satisfying **A5** but such that $B \neq \mathbf{1}$.

We now consider another of Lyndon's axioms:

A0. $x \neq \mathbf{1}$ implies that $|x| < |x^2|$.

It follows immediately from Lemma 3·1 and Theorem 8·1 of (4) that **A0** places a rather large restriction on an NFS-group, namely:

LEMMA 3·7. *An NFS-group whose length function satisfies* **A0** *has trivial base group and is a free group.*

Note that Lemma 3·7 does not say that this length function is the usual length function on a free group.

Example. Let $G = \langle a, b \rangle$, $B = \mathbf{1}$ and $X = \{a, b, ab\}$, then (G, B, X) is an NF-structure on G satisfying **A0**, but its length function is clearly not the usual free group length function.

I would like to thank the referee for suggesting many detailed improvements to this paper.

REFERENCES

(1) CHISWELL, I. M. Abstract length functions in groups. *Math. Proc. Cambridge Philos. Soc.* **86** (1976), 451–463.

(2) HOARE, A. H. M. On length functions and Nielson methods in free groups. *J. London Math. Soc.* **14** (1) (1976), 188–192.

(3) HURLEY, B. M. Embedding theorems for groups, M.Sc. dissertation (Oxford, 1972).

(4) LYNDON, R. C. Length functions in groups. *Math. Scand.* **12** (1963), 209–234.

https://doi.org/10.1017/S0305004100055286 Published online by Cambridge University Press