1 Uzupełnienie o wolnej algebrze Liego

Wiemy że dla wolnej algebry Liego F generowanej przez zbiór X uniwersalna algebra obwiednia $U(F)$ to algebra tensorowa $T(M)$ generowane przez moduł wolny generowany przez X. Wiemy że jeśli F jest modułem wolnym to wkłada się różnowartościowo w $U(F)$. Jednakże na razie wiemy że F jest modułem wolnym tylko wtedy gdy pierścień podstawowy R jest ciałem. Chcemy pokazać że F jest modułem wolnym w ogólnym przypadku. Wystarczy to zrobić dla $R = \mathbb{Z}$. Mianowicie dla ogólnego R mamy

$$F_R = R \otimes_{\mathbb{Z}} F_{\mathbb{Z}}$$

gdzie dla F_R i $F_{\mathbb{Z}}$ indeks oznacza że rozważamy algebrę nad danym pierścieniem, zaś $\otimes_{\mathbb{Z}}$ oznacza że moduły traktujemy jako moduły nad $\mathbb{Z}$. Łatwo pokazać że jeśli $F_{\mathbb{Z}}$ jest modułem wolnym nad $\mathbb{Z}$ to produkt wyżej jest modułem wolnym nad R.

Aby pokazać wynik dla $\mathbb{Z}$ podamy alternatywną konstrukcję F.

Definicja. Magmą nazywamy zbiór z działaniem dwuargumentowym o którym nie robimy żadnych dodatkowych założeń.

Wolna magma generowana przez X to magma S z odwzorowaniem $\iota : X \to S$ taka że dla każdej magmy V i odwzorowania $f : X \to V$ istnieje dokładnie jeden homomorfizm h z S w V taki że $f = h \circ \iota$.

Wolną magmę łatwo zbudować: bierzemy drzewa binarne takie że liśćmi są elementy X. Dwa drzewa a i b mnożywmy w ten sposób że budujemy drzewo którego lewym poddrzewem jest a zaś przawym poddrzewem jest b. Oczywiście znając obrazy X-ów możemy obliczyć rekursywnie obraz drzewa: najpierw obliczamy obrazy poddrzew, a potem wymnażamy otrzymane wartości. Widać że jest to homomorfizm i że jest to jedyne przedłużenie f które daje homomorfizm. A więc jest spełniona własność uniwersalna, czyli otrzywmaliśmy wolną magmę.

Uwaga: o elementach wolnej magmy S możemy myśleć jako o wyrażeniach (drzewach wyrażeń). Alternatywnie, możemy myśleć że elementy S to ciągi elementów X uzupełnione nawiasami tak by kolejność działań była jednoznaczna.

Mając wolną magmę S rozpatrujemy wolną algebrą niełączną N nad $\mathbb{Z}$, tzn. moduł wolny nad $\mathbb{Z}$ z bazą S. W N elementy bazy mnożymy zgodnie z regułami S i przedłużamy działanie $\mathbb{Z}$-liniowo na N. Łatowo zobaczyć że N spełnia następującą własność uniwersalną: dla dowolnego odwzorowania $f : X \to A$ gdzie A jest algebrą niełączną nad $\mathbb{Z}$ istnieje dokładnie jeden homomorfizm algebr niełącznych h z N w A taki że $f = h \circ \iota$ gdzie ι jest włożeniem X w N. Mianowicie, dostajemy homomorfizm magm z S w A. Ten homomorfizm rozszerza się $\mathbb{Z}$-liniowo na N.

W N rozpatrujemy ideał I (tzn. podmoduł N zamknięty na mnożenie z lewej i prawej strony przez elementy N) generowany przez elementy postaci $s \cdot s$ i $s \cdot (t \cdot u) - (s \cdot t) \cdot u - t \cdot (s \cdot u)$. Dzielenie przez ideał I oznacza że w N/I jest spełniona antysymetria i tożsamość Jacobiego, czyli N/I jest algebrą Liego.

Twierdzimy że N/I ma własność uniwersalną wolnej algebry Liego nad $\mathbb{Z}$. Mianowicie, jeśli A jest algebrą Liego, to jest też algebrą niełączną, czyli dostajemy homomorfizm h z N w A. Widać że elementy I są w przeprowadzane na 0 w A, czyli I jest w jądrze h, czyli h daje hpmomorfizm algebry ilorazowej N/I w A. Jedność otrzymanego homomorfizmu jest oczywista.

Zauważmy teraz że w N można wprowadzić pojęcie stopnia elementu: stopień elementu S to ilość liści drzewa, stopień elementu $n \in N$ to maksimum

stopni $s \in S$ w przedstawieniu n jako kombinacji liniowej elementów S z niezerowymi współczynnikami. Widać że stopień iloczynu to iloczyn stopni. Element $n \in N$ nazywamy jednorodnym jeśli wszystkie elementy $s \in S$ z niezerowymi współczynnikami w przedstawieniu n mają ten sam stopień. Możemy zapisać

$$N = \oplus_{k=1}^{\infty} N_k$$

gdzie N_k to zbiór elementów N jednorodnych stopnia k.

Mamy też

$$I = \oplus_{k=1}^{\infty} I_k$$

gdzie I_k to zbiór elementów I jednorodnych stopnia k. Ta ostatnia równość wynika z tego że jako generatory I można wybrać elementy jednorodne. Dokładniej, tożsamość Jacobiego jest trójliniowa, więc mając ją dla elementów jednorodnych przez wieloliniowość otrzymamy ją dla sum. Równość $s \cdot s = 0$ jest nieco bardziej kłopotliwa. Zauważmy że implikuje ona antysymetrię $s \cdot t + t \cdot s = 0$. Antysymetria jest dwuliniowa, więc z elementów jednorodnych rozszerzy się na dowolne. Teraz piszemy $s = \sum_{i=1}^{k} s_i$ gdzie s_i jest jednorodny stopnia i. Mamy

$$s \cdot s = \sum_{i=1}^{k} s_i \cdot s_i + \sum_{j<i} s_j \cdot s_i + s_i \cdot s_j.$$

Druga suma zniknie na mocy antysymetrii, a w pierwszej sumie $s_i \cdot s_i = 0$ jest jednorodne. A więc faktycznie I jest generowany przez elementy jednorodne.

Mamy teraz

$$N/I = \oplus_{k=1}^{\infty} N_k/I_k.$$

Innymi słowy, N/I jest generowane przez elementy jednorodne. Zauważmy, że jeśli X jest zbiorem skończonym to N_k/I_k jest modułem skończenie generowalnym.

Lemat 1.1 *Jeśli R jest ciałem a X jest zbiorem skończonym, to wymiar $R \otimes_{\mathbb{Z}} (N_k/I_k)$ nie zależy od R.*

Dowód. Wiemy że $R \otimes_{\mathbb{Z}} (N/I) = F_R$ wkłada się w algebrę tensorową $T(M)$ gdzie M jest modułem wolnym (przestrzenią wektorową) z bazą X. Przy tym $R \otimes_{\mathbb{Z}} (N_k/I_k)$ odwzorowuje się wzajemnie jednoznacznie na elementy jednordne stopnia k w F. A więc wystarczy pokazać że wymiar przestrzeni elementów jednorodnych stopnia k w F nie zależy od R. Jednakże, na mocy lematu o bazie algebry obwiedniej (twierdzenia PBW) przy ustalonym porządku na bazie F elementy postaci

$$e_1 \dots e_k$$

z $e_1 \leq e_2 \leq \dots \leq e_k$ gdzie e_i są elementami bazy F stanowią bazę $U(F) = T(M)$. Jeśli bazę F wybierzemy z elementów jednorodnych, to otrzymane elementy bazy $T(M)$ też będą jednorodne. A więc produkty wyżej stopnia k dają nam bazę $M^{\otimes k}$. $M^{\otimes k}$ ma wymiar $|X|^k$. Zauważmy teraz że możemy rekursywnie wyliczać wymiar przestrzeni rozpinanej przez elementy jednorodne F stopnia k. Mianowicie, $M^{\otimes k}$ jest suma prostą podprzestrzeni rozpinanej przez elementy bazy F rzędu k i podprzestrzeni rozpinanej przez produkty elementów bazy F niższego stopnia. Z założenia indukcyjnego ilość (elementów bazy F niższego stopnia (dla każdego stopina z osobna) jest znana. To pozwala wyznaczyć

wymiar przestrzeni rozpinanej przez produkty elementów bazy F. Teraz ilość elementów jednorodnych bazy F stopnia k otrzymujemy jako różnicę wymiaru $M^{\otimes k}$ (czyli $|X|^k$) i ilości produktów elemntów bazy F niższych stopni. $\square$

Uwaga: Rozwijając powyższy argument można pokazać że

$$\dim(R \otimes_{\mathbb{Z}} (N_k/I_k)) = \frac{1}{k} \sum_{m|k} \mu(m) d^{n/m}$$

gdzie $\mu(m)$ jest funkcją Möbiusa, tzn. $\mu(m) = 0$ jeśli m jest podzielne przez kwardrat liczby pierwszej, w przeciwnym razie $\mu(m) = (-1)^l$ gdzie l jest ilością czynników pierwszych m.

Lemat 1.2 *Jeśli M jest skończenie generowalnym modułem nad $\mathbb{Z}$ i wymiar $\mathbb{Z}_q \otimes_{\mathbb{Z}} M$ gdzie $\mathbb{Z}_q$ jest ciałem reszt modulo q traktowanym jako $\mathbb{Z}$-moduł nie zależy od liczby pierwszej q to M jest modułem wolnym.*

Dowód. Na mocy twierdzenia o strukturze skończenie generowanych grup abelowych M można zapisać jako

$$M = \mathbb{Z}^k \oplus \oplus_p M_p$$

gdzie M_p jest podmodułem elementów p-torsyjnych, tzn. istnieje l takie że $p^l M_p = \{0\}$. Przy tym tylko skończenie wiele składników w sume wyżej jest niezerowe. Zauważmy teraz że jeśli q jest liczbą pierwszą różną of p to p^l jest odwracalne w $\mathbb{Z}_q$, czyli

$$\mathbb{Z}_q \otimes_{\mathbb{Z}} M_p = \{0\}.$$

Jeśli M_p jest nietrywialne i $q = p$ to również $\mathbb{Z}_q \otimes_{\mathbb{Z}} M_p$ jest nietrywialne. Wybierając q różne od p z nietrywialnym M_p widzimy że

$$\mathbb{Z}_q \otimes_{\mathbb{Z}} M = \mathbb{Z}_q^k$$

czyli wymiar jest równy k. Gdyby któreś z M_p było nietrywialne, to biorąc $q = p$ otrzymalibyśmy wymiar większy niż k. A więc niezależność wymiaru od q implikuje że wszystkie M_p są trywialne, co z kolei oznacza że M jest modułem wolnym. $\square$

Lemat 1.3 *Wolna algebra Liego F_R jest modułem wolnym dla dowolnego pierścienia R i dowolnego zbioru generatorów X.*

Dowód. Jeśli $R = \mathbb{Z}$ i X jest zbiorem skończonym to Lemat 1.1 i Lemat 1.2 implikują że

$$N_k/I_k$$

jest modułem wolnym nad $\mathbb{Z}$, a więc również F jest modułem wolnym jako suma prosta modułów wolnych.

Rozważmy teraz nieskończone X. Wystarczy pokazać że naturalne odwzorowanie z F w $T(M)$ jest różnowartościowe. Mianowicie, wtedy F jest izomorficzne z podmodułem $T(M)$. Lecz $T(M)$ jest modułem wolnym a nad $\mathbb{Z}$

podmoduł modułu wolnego jest wolny, więc również F jest modułem wolnym. Przypuśćmy że $x \in F$ jest niezerowy, ale ma zerowy obraz w $T(M)$. x zawiera tylko skończenie wiele zmiennych, czyli jest w obrazie pewnego N_{X_0} gdzie X_0 jest skończonym podzbiorem X zaś N_{X_0} oznacza wolną algebrę nieprzemienną generowaną przez X_0. x zadaje więc element $\tilde{x}$ wolnej algebry Liego $\tilde{F}$ generowanej przez X_0. x jest obrazem $\tilde{x}$ przez naturalne włożenie $\tilde{F}$ w F, a więc $\tilde{x}$ jest niezerowy. Skoro X_0 jest skończony, to $\tilde{F}$ jest modułem wolnym nad $\mathbb{Z}$, a więc wkłada się różnowartościowo w $U(\tilde{F})$ a więc również w $T(M)$. Daje to sprzeczność z przypuszczeniem że obraz x jest zerem, co pokazuje że naturalne odwzorowanie z F w $T(M)$ faktycznie jest różnowartościowe, czyli F jest modułem wolnym.

Jak zauważyliśmy na początku tej części, wynik dla $\mathbb{Z}$ implikuje wynik dla dowolnego pierścienia. □

2 Bazy Halla

Nasze dotychczasowe wyniki są zadowalające z teoretycznego punktu widzenia i teoretycznie wystarczają do prowadzenia obliczeń w wolnej algebrze Liego. Jednakże można uzyskać nieco bardziej jawny opis bazy wolnej algebry Liego. Taki opis znacząco ułatwia obliczenia. Może też dać dodatkowe intuicje o budowie wolnej algebry Liego.

Przy budowie bazy Halla musimy pracować z trzema różnymi zbiorami: elementami wolnej magmy S, elementami wolnej algebry Liego F będącymi obrazami elementów S i obrazami elementów S w półgrupie wolnej generowanej przez X. Ogólnie jedno słowo w półgrupie wolnej jako przeciwobrazy ma wiele elementów S. Jednakże w konstrukcji Halla są dodatkowe warunki które powodują że dane słowo ma tylko jeden przeciwobraz spełniający te warunki. O konstrukcji Halla można myśleć jako o rozmieszczeniu nawiasów w słowie, po dodaniu nawiasów dostaniemy element (lub ciąg elementów) wolnej magmy S które można odwzorować w F. Dodajmy że właściwe operacje to te z F, ale większość rozumowań prowadzi się na elementach S albo słowach.

Istotną częścią konstrukcji Halla są przekształcenia otrzymanych elementów. Te przekształcenia zależą od porządku na tych elementach. W konstrukcji można używać wiele różnych porządków i zmieniając porządek można uzyskać trochę dodatkowych wyników. Jednakże można by też się ograniczyć do jednego porządku. Przykładowy porządek który działa to porządek leksykograficzny na słowach, gdzie startujemy z porządku liniowego na S i mówimy że słowo a jest większe od słowa b jeśli pierwsza litera licząc od prawej w a jest większa od pierwszej litery b. W przypadku gdy b jest prefiksem a to uznajemy a za większe.

Uwaga: Porządek leksykograficzny na słowach nie wyznacza porządku liniowego na S. Jednakże dalej uzasadnimy że porządek leksykograficzny na słowach wystarcza do konstrukcji którą przedstawimy.

Uwaga: Zwykle porządek leksykograficzny na słowach definiuje się inaczej, zaczynając od lewej strony. Większość wyników by wtedy dalej działała, ale lemat charakteryzujący słowa Halla względem porządku leksykograficznego zależy od detali porządku. Nieco ogólniej, są możliwe różne konwencje i zależnie od

wyboru konwencji pewne fragmenty są mniej lub bardziej naturalne. Dziwna definicja porządku leksykograficznego wydaje się być małym kosztem w stosunku do bardziej naturalnego przedstawienia w innych miejscach.

Potrzebujemy jeszcze trochę notacji: gdy $x = a \cdot b$ jest elementem S to piszemy $a = L(x)$, $b = R(x)$. Innymi słowy L daje lewe poddrzewo, R daje prawe poddrzewo (te operacje nie są zdefiniowane gdy x jest zmienną). Dla elementów $s \in S$ jak poprzednio definiujemy stopień $\deg(s)$ jako długość odpowiadającego mu słowa.

Definicja. Powiemy że podzbiór $H \subset S$ z liniowym porządkiem jest zbiorem Halla jeśli spełnia następujące warunki

- $a \cdot b \in H$ wtedy i tylko wtedy gdy $a, b \in H$, $a > b$ i albo $a \in X$ lub $a = c \cdot d$ i $d \leq b$,
- dla $a \cdot b \in H$ mamy $a \cdot b > b$,
- $X \subset H$.

Lemat 2.1 *Obraz H w F generuje F jako R moduł.*

Dowód. Bez utraty ogólności można zakładać że X jest skończony. Z definicji $X \subset H$, czyli H generuje F jako algebrę Liego. Aby pokazać lemat wystarczy więc pokazać że nawias Liego obrazów elementów H jest $\mathbb{Z}$-liniową kombinacją obrazów elementów H. W dowodzie oznaczymy przez f naturalne odwzorowanie z S w F. Dowód będzie indukcyjny, by indukcja działała potrzebujmy nieco mocniejszy warunek

$$[f(a), f(b)] = \sum_i c_i f(d_i)$$

z $c_i \in \mathbb{Z}$, i $d_i > \min(a, b)$. Ponadto żądamy by stopień $\deg(d_i) = \deg(a) + \deg(b)$. Indukcja jest ze względu na $\deg(a) + \deg(b)$, a w przypadku równości stopni ze względu na odwrócone $\min(a, b)$, tzn. zakładamy że wynik zachodzi dla mniejszej sumy długości i dla równej sumy długości z większym $\min(a, b)$. Zakładamy że X jest skończony, więc jest tylko skończenie wiele drzew danej długości, więc tylko skończenie wiele możliwości na $\min(a, b)$. A więc będzie działała zasada indukcji. Ze względu na antysymetrię możemy zmienić kolejność a i b, co pozwala zakładać że $a > b$. Jeśli wtedy $a \in X$ lub $a = e \cdot g$ i $g \leq b$, to $a \cdot b \in H$ i $[f(a), f(b)] = f(a \cdot b)$. Ponadto $a \cdot b > b = \min(a, b)$. A więc pozostaje rozważyć przypadek gdy $a = e \cdot g$ i $g > b$. Wtedy na mocy własności porządku na zbiorze Halla mamy $e > g > b$. Na mocy tożsamości Jacobiego mamy

$$[f(a), f(b)] = [[f(e), f(g)], f(b)] = [f(e), [f(g), f(b)]] + [[f(e), f(b)], f(g)].$$

Na mocy założenia indukcyjnego

$$[f(g), f(b)] = \sum r_i f(v_i),$$

gdzie r_i są liczbami całkowitymi, $\deg(v_i) = \deg(g) + \deg(b)$ i $v_i > \min(g, b) = b$. A więc $\deg(e) + \deg(v_i) = \deg(e) + \deg(g) + \deg(b) = \deg(a) + \deg(b)$, czyli dla par e, v_i suma stopni się nie zmienia a $\min(e, v_i) > b = \min(a, b)$. Czyli do

$$[f(e), [f(g), f(b)]]$$

stosuje się założenie indukcyjne i możemy to zapisać jako $\mathbb{Z}$-liniową kobinację obrazów elementów H z tą samą sumą stopni i elementami $> \min(a, b)$.

Podobnie

$$[f(e), f(b)] = \sum s_i f(w_i)$$

gdzie s_i są liczbami całkowitymi, $\deg(w_i) = \deg(e) + \deg(b)$, i $w_i > \min(e, b) = b$. Znowu suma stopni się zgadza, zaś $\min(w_i, g) > b$, czyli również w tym przypadku stosuje się założenie indukcyjne. □

Lemat 2.2 *Oznaczmy przez g naturalne odwzorowanie z S w półgrupę wolną i niech H będzie zbiorem Halla. Niech w będzie elementem półgrupy wolnej (słowem). Wtedy istnieje dokładnie jeden ciąg niemalejący $h_i \in H$, $i = 1..l$ taki że*

$$w = g(h_1)g(h_2)\dots g(h_l)$$

Aby udowodnić ten lemat potrzebujmy dodatkowe defincje i lematy pomocnicze.

Definicja. Powiemy że ciąg elementów $h_i \in H$ jest dopuszczalny jeśli dla każdego i albo $h_i \in X$, albo $R(h_i) \leq h_j$ dla $j > i$.

Oczywiście ciąg elementów X jest dopuszczalny. Również niemalejący ciąg elementów H jest dopuszczalny. Mianowicie, z własności porządku na zbiorze Halla $R(h_i) < h_i$, czyli skoro ciąg jet niemalejący to $R(h_i) < h_i \leq h_j$ dla $j < i$.

W dalszym ciągu potrzebne nam będą przekształcenia ciągów dopuszczalnych (system przepisywania). Gdy a i b są ciągami dopuszczalnymi to piszemy $a \to b$ (a może być przepisany w jednym kroku do b) gdy b różni się od a tym że dwa sąsiednie elementy a są zastąpione przez ich produkt w S. $\to$ jest relacją (może być więcej niż jeden wybór pary która da ciąg dopuszczalny. Powiemy że a jest nieredukowalny jeśli nie istnieje b taki że $a \to b$. Relacja $\to^*$ jest domknięciem tranzytywnym $\to$, tzn. $a \to^* a$ i jeśli $a \to b$ oraz $b \to^* c$ to $a \to^* c$. Przy tym $a \to^* b$ wtedy i tylko wtedy gdy wynika to z podanych reguł.

Zauważmy że ciągi niemalejące są nieredukowalne: gdy $a_i \leq a_{i+1}$ to produkt $a_i \cdot a_{i+1}$ nie należy do zbioru Halla. Jeśli $a \to b$ to b ma mniejszą długość niż a, a więc po skończonej liczbie kroków otrzymamy ciąg nieredukowalny.

Lemat 2.3 *Niech g oznacza naturalne odwzorowanie z S w półgrupę wolną generowaną przez X (czyli słowa z lierami z X) i niech H będzie zbiorem Halla w S. Jeśli $h \to^* t$ to*

$$g(h_1)g(h_2)\dots g(h_k) = g(t_1)g(t_2)\dots g(t_l).$$

Dla danego ciągu dopuszczalnego $h_i \in H$, $i = 1, \dots k$ istnieje niemalejący ciąg $t_i \in H$, $i = 1, \dots, l$ taki że $h \to^* t$. Ponadto ciągi nieredukowalne są niemalejące.

Dowód. Gdy $h \to t$ to równość

$$g(h_1)g(h_2)\dots g(h_k) = g(t_1)g(t_2)\dots g(t_l)$$

jest oczywista. Jeśli $h \to^* t$ to równość wyżej otrzymujemy przez indukcję.

Pozostaje pokazać istnienie niemalejącego t takiego że $h \to^* t$. Dowód jest indukcyjny ze względu na długość h. Jeśli ciąg h_i jest niemalejący, to mamy wynik. W przeciwnym razie istnieje i takie że $h_i > h_{i+1}$. Biorąc największe takie i mamy $h_{i+1} \leq h_{i+2} \leq \cdots \leq h_k$. Teraz zastępujemy parę h_i i h_{i+1} przez $e = h_i \cdot h_{i+1}$. Mamy $h_i > h_{i+1}$ i $R(h_i) \leq h_{i+1}$, a więc $e \in H$. Ponadto $R(e) = h_{i+1} \leq h_j$ dla $j > i+1$, więc warunek dopuszczalności jest spełniony dla e. Ciąg h_j z $j > i+1$ jest jak poprzednio, więc warunek dopuszczalności jest spełniony dla tych elementów. Dla h_j z $j < i$ jest zmiana, musimy dodatkowo porównać $R(h_j)$ z e. Lecz $e > h_{i+1} \geq R(h_j)$ więc nowy ciąg jest dopuszczalny. Zmniejszyliśmy długość ciągu, więc po skończonej ilości kroków proces się zatrzyma i otrzymamy ciąg niemalejący. □

Lemat 2.4 *Oznaczmy przez g naturalne odwzorowanie z S w półgrupę wolną i niech H będzie zbiorem Halla. Niech w będzie elementem półgrupy wolnej (słowem). Niech $\tilde{w}$ będzie ciągiem liter w. Jest to ciąg dopuszczalny. Niech h z $h_i \in H$, $i = 1..l$ będzie ciągiem dopuszczalnym takim że*

$$w = g(h_1)g(h_2)\ldots g(h_l)$$

Twierdzimy że $\tilde{w} \to^* h$.

Dowód: Dowód indukcyjny, ze względu na odwrotny porządek dla długości. Tzn. możemy zakładać wynik dla ciągów dłuższych niż l i potrzebujemy pokazać dla l. Jeśli h składa się tylko z liter to $\tilde{w} = h$ czyli $\tilde{w} \to^* h$. W przeciwnym razie niech i będzie najmniejszym i takim że h_i nie jest literą, czyli $h_i = a \cdot b$. Ciąg t budujemy zastępując w h element h_i przez parę elementów a i b. Dla $j < i$ elementy h_j są literami więc warunek dopuszczalności jest spełniony. $h_i \in H$ więc $R(a) \leq b = R(h_i) \leq h_j$ dla $j > i$, czyli warunek dopuszczalność jest spełniony dla a. Na mocy własności porządku $R(b) < b = R(h_i)$ czyli warunek dopuszczalności będzie spełniony dla b. Końcowy ciąg elementów t zaczynając po b jest taki sam jak w h, więc warunek dopuszczalności będzie spełniony dla tych elementów. Oczywiście $t \to h$ i t jest dłuższy niż h, więc do t stosuje się założenie indukcyjne i mamy $\tilde{w} \to^* t$. Ale wtedy $\tilde{w} \to^* h$ □

Lemat 2.5 *Jeśli a, b, c są ciągami dopuszczalnymi takimi że $a \to b$, $a \to c$ i $b \neq c$ to istnieje ciąg dopuszczalny d taki że $b \to d$ i $c \to d$.*

Dowód: Niech $b_i = a_i \cdot a_{i+1}$ i $c_j = a_j \cdot a_{j+1}$. Zamieniając miejscami b i c można zakładać że $i < j$. Gdyby $i = j + 1$ to mielibyśmy $a_i > a_{i+1} > a_{i+1}$. Lecz $R(a_i \cdot a_{i+1}) = a_{i+1} > a_{i+2}$ co przeczyłoby dopuszczalności b. A więc $j \geq i + 2$ (zmiany się nie nakładają). Niech d będzie ciągiem otrzymanym z b zastępując parę $b_{j-1} = a_j$ i $b_j = a_{j+1}$ przez $b_{j-1} \cdot b_j = a_j \cdot a_{j+1}$. Ciąg ten jest dopuszczalny. Mianowicie końcowy ciąg elementów d zaczynając od d_{j-1} jest taki sam jak końcowy ciąg elementów c zaczynając od c_j. Skoro ciąg c jest dopuszczalny to na pozycjach $\geq j - 1$ warunek dopuszczalności jest spełniony dla d. Na pozycjach $< j - 1$ przy sprawdzaniu dopuszczalności różnica w porównaniu do b jest taka że zastąpiliśmy parę b_{j-1} i b_j przez $b_{j-1} \cdot b_j$. Lecz $b_{j-1} \cdot b_j > b_j$, więc warunek dopuszczalności będzie spełniony. Oczywiście $b \to d$.

Ciąg d można też otrzymać z c zastępując $c_i = a_i$ i $c_{i+1} = a_i$ przez $c_i \cdot c_{i+1} = a_i \cdot a_{i+1}$. Dopuszczalnośc d już pokazaliśmy, więc $c \to d$. □

Lemat 2.6 *Jeśli ciągi a, b, h są dopuszczalne, h jest niemalejący, $a \to b$, $a \to^* h$ to $b \to^* h$.*

Dowód: Indukcja ze względu na ilość aplikacji $\to$ w $a \to^* h$. a jest redukowalne, h jest nieredukowalne więc istnieje ciąg dopuszczalny c taki że $a \to c$ i $c \to^* h$. Jeśli $c = b$ to kończy dowód. W przeciwnym razie, na mocy Lematu 2.5 istnieje ciąg dopuszczalny d taki że $b \to d$ i $c \to d$. W $c \to^* h$ ilość aplikacji $\to$ jest zmniejszona o 1, więc na mocy założenia indukcyjnego $d \to^* h$. Lecz wtedy też $b \to^* h$. □

Dowód lematu 2.2: Ciąg zmiennych jest dopuszczalny, więc na mocy Lematu 2.3 dla każdego w istnieje ciąg niemalejący h taki że $\tilde{w} \to^* h$ i

$$w = g(h_1)g(h_2)\dots g(h_k)$$

co dowodzi istnienia. By pokazać jednoznaczność zauważmy że jeśli

$$w = g(h_1)g(h_2)\dots g(h_k)$$

i ciąg h jest niemalejący to na mocy Lematu 2.4 mamy $\tilde{w} \to^* h$. A więc wystarczy pokazać że $a \to^* h$ i $a \to^* t$ dla niemalejących ciągów h i t i dopuszczalnego a implikuje $h = t$. Robimy to indukcyjnie ze względu na ilość aplikacji $\to$ w $a \to^* t$. Jeśli jest to 0, tzn. $a = t$, to a jest nieredukowalny i $a = h$. W przeciwnym razie, niech $a \to b$ i $b \to^* t$. Na mocy Lematu 2.6 wtedy $b \to^* h$. W $b \to^* t$ ilość aplikacji $\to$ jest zmniejszona o 1, więc stosuje się założenie indukcyjne i mamy $h = t$. □

Lemat 2.7 *Niech H będzie zbiorem Halla w S. Wtedy obraz w F przez odwzorowanie naturalne jest bazą algebry wolnej.*

Dowód: Na mocy Lematu 2.1 obraz H generuje F. Pozostaje pokazać liniową niezależność nad $\mathbb{Z}$. Wystarczy to zrobić dla skończonego zbioru X. Dla skończonego X rozważamy $F_{\mathbb{Q}} = \mathbb{Q} \otimes_{\mathbb{Z}} F$, i w $F_{\mathbb{Q}}$ elementy jednorodne stopnia k. Na mocy Lematu 2.2 ilość elementów H stopnia k jest równa wymiarowi F_k. Dokładniej, oba ciągi liczb spełniają tą samą rekurencję która pozwala jednoznacznie wyznaczyć te liczby (jest to powtórzenie rozumowania z dowodu Lematu 1.1). Gdyby elementy H były liniowo zależne, to wymiar byłby miejszy, co dałoby sprzeczność. □

Mówmy że słowo w jest słowem Halla jeśli $w = g(h)$, $h \in H$ zaś g jest naturalnym odwzorowaniem z S w półgrupę wolną. Na mocy Lematu 2.2 h jest jednoznacznie wyznaczone przez w, czyli porządek na H jest wyznaczony przez

porządek na słowach Halla. Jednakże, konstrukcja zbioru Halla i dowodu jego własności (takich jak Lematu 2.2) używa porządek, więc jest ryzyko błędnego koła. Jednakże można rozumować następująco: indukcyjnie budujemy zbiory H_k i porządek na H_k rozszerzający porządek na słowach. $H_1 = X$. H_{k+1} zawiera H_k i dodane elementy. Element $a \cdot b$, $a, b \in H_k$ dodajemy do H_{k+1} wtedy i tylko wtedy gdy $a \cdot b > b$ i są spełnione warunki niżej. Po pierwsze zgodnie z porządkiem na słowach $g(a \cdot b) > g(b)$ (równość nie jest możliwa, bo $g(a \cdot b)$ jest dłuższe niż $g(b)$, czyli $g(a \cdot b) \neq g(b)$). Po drugie albo $a \in H_1$ $a = c \cdot d$ i $d \leq b$. Zauważmy że porównania używają już zdefiniowany porządek na H_k. Na H_{k+1} mamy porządek częściowy, zadany przez porządek na H_k i porządek na słowach. To jest faktycznie porządek częściowy, bo porządek na H_k rozszerza porządek na słowach. Teraz rozszerzamy porządek na H_{k+1} do porządku liniowego. Suma

$$\bigcup_{k=1}^{\infty} H_k$$

jest oczywiście zbiorem Halla. Do tego zbioru stosuje się Lemat 2.2 pokazując że porządek jest jednoznacznie wyznaczony przez porządek na słowach Halla.

Zanotujmy prostą własność porządku leksykograficznego: jeśli b nie jest sufiksem a i $a > b$ to dla dowolnych c i d mamy $ca < db$. Mianowicie, wynik porównania $a > b$ jest wyznaczony przez porównanie liter, startując od końca i któraś z liter da nam wynik (nie jest możliwy przypadek równości liter, bo wtedy albo a byłoby sufiksem b, co by dało $b \geq a$, lub b byłoby sufiksem a, co wykluczamy z założenia). Oznacza to że dopisywanie dodatkowych liter na początku nie zmienia wyniku porównania. W szczególności ta własność zachodzi gdy długość a jest mniejsza lub równa długości b i $a > b$ (przy równej długości a będąc sufiksem byłoby równe b, co jest sprzeczne z $a > b$).

Lemat 2.8 *Słowo w jest słowem Halla względem porządku leksykograficznego wtedy i tylko wtedy gdy jest mniejsze od dowolnego właściwego prefiksu w.*

Dowód: Najpierw indukcyjnie pokażemy że słowa Halla spełniają podany warunek. Gdy w jest literą to $w \in H$ i zbiór prefiksów właściwych w jest pusty, czyli warunek jest trywialnie spełniony. W kroku indukcyjnym możemy zakładać że wynik zachodzi dla krótszych słów. Jeśli $w = g(a \cdot b)$ gdzie $a, b \in H$ to niech k, l będą długościami a i odpowiednio b. Jeśli v jest prefiksem w którego długość m jest większa niż k, to ciąg u końcowych $m - k$ liter v jest prefiksem $g(b)$. $g(b)$ jest słowem Halla, więc na mocy założenia indukcyjnego $u > g(b)$. Długość u jest mniejsza niż długość $g(b)$, więc na mocy własności wyżej mamy $v > w$. Jeśli v jest prefiksem którego długość jest $= k$ to $v = g(a) > g(b)$. Jeśli $g(b)$ nie jest sufiksem v, to dalej na mocy własności wyżej $v < w$. Jeśli $g(b)$ jest sufiksem v, tzn. $v = ug(b)$ to wynik porównania nie zależy od ostanich l liter, czyli dostaniemy go porównując u z $v = g(a)$. Na mocy założenia indukcyjnego $u > v$, czyli również $v > w$. Jeśli długość v jest $< k$, to v jest właściwym prefiksem $g(a)$, czyli z założenia indukcyjnego $v > g(a)$. Ale pokazaliśmy już że $g(a) > w$, czyli $v > w$, co kończy dowód tej części.

W drugą stronę załóżmy że każdy właściwy prefix w jest większy niż w. Na mocy Lematu 2.2 słowo w można zapisać jako produkt słów Halla h_i:

$$w = h_1 h_2 \dots h_l$$

z $h_1 \leq h_2 \leq \cdots \leq h_l$. Jeśli $l = 1$ to w jest słowem Halla co daje wynik. Jeśli $l > 1$, $h_1 = h_l$ to $h_1 < w$ co daje sprzeczność z założeniem. A więc można zakładać że $h_1 < h_l$. Jeśli h_1 nie jest sufiksem h_l to na mocy własności wyżej mamy $h_1 < w$ co znowu daje sprzeczność z założeniem. Jeśli h_1 jest sufiksem h_l, to jest też sufiksem w, czyli skoro w jest dłuższe, to $h_1 < w$. A więc przypuszczenie że $l > 1$ w każdym przypadku prowadzi do sprzeczności. □