Word maps with constants on symmetric groups

Jakob Schneider  |  Andreas Thom

Fakultät Mathematik, TU Dresden, Dresden, Germany

Correspondence
Jakob Schneider, Institut für Geometrie, Fakultät Mathematik, TU Dresden, 01062 Dresden, Germany.
Email: jakob.schneider@tu-dresden.de

To Doris, Hannah, Torsten, Jonathan, and Paul who supported me during the long time of my sickness. J. Schneider.

Funding information
ERC Consolidator Grant, Grant/Award Number: 681207

Abstract
We study word maps with constants on symmetric groups. Even though there are non-trivial mixed identities of bounded length that are valid for all symmetric groups, we show that no such identities can hold in the limit in a metric sense. Moreover, we prove that word maps with constants and non-trivial content, that are short enough, have an image of positive diameter, measured in the normalized Hamming metric, which is bounded from below in terms of the word length. Finally, we also show that every self-map \( G \to G \) on a finite non-abelian simple group is actually a word map with constants from \( G \).

KEYWORDS
Hamming metric, mixed identities, symmetric groups, word maps, word image, word length, words with constants

MSC (2020)
20D06, 20F69

1  |  INTRODUCTION

Recently, there has been increasing interest in word maps and laws on finite, algebraic, and topological groups [2, 3, 6–12, 14, 16, 18–22, 24, 29, 32]. Here, every word \( w \in F_r = \langle x_1, \ldots, x_r \rangle \) induces a word map \( w : G^r \to G \) on every group \( G \) by substitution, where \( F_r \) denotes the free group in the \( r \) generators \( x_1, \ldots, x_r \) and we set \( w(g_1, \ldots, g_r) \) to be the image of \( w \) under the unique homomorphism \( F_r \to G \) which maps \( x_i \mapsto g_i \) (\( i = 1, \ldots, r \)). The word \( w \) is called an identity or a law for \( G \) iff \( w(g_1, \ldots, g_r) = 1_G \) for all choices of the \( g_i \in G \) (\( i = 1, \ldots, r \)). It is an interesting question to study the length of the shortest non-trivial law of a given finite group \( G \). This was done in [4, 31]. A bit less restrictively, one can ask when the image of a word map is small in a metric sense, see, for example, [18, 24, 30].

In this paper, we study word maps with constants. A word with constants in \( G \) is an element of the free product \( F_r \ast G \). We get an associated map \( G^n \to G \) in a similar way as before by replacing the variables with elements from \( G \). The word is called a mixed identity or a law with constants iff \( w(g_1, \ldots, g_r) = 1_G \) for all choices of the \( g_i \in G \) (\( i = 1, \ldots, r \)).

Following [16, 27], for \( w \in F_r \ast G \), we study the augmentation \( \varepsilon(w) \in F_r \), which replaces all constants by the neutral element. We call \( \varepsilon(w) \in F_r \) the content of \( w \). It has been observed in different circumstances [16, 27] that word maps with constants which have a non-trivial content, tend to have a large image. We prove a corresponding result for symmetric groups with respect to the normalized Hamming metric. For \( \sigma \in S_n \), we write

\[
\text{fix}(\sigma) := \{ x \mid x.\sigma = x \}
\]
for the set of fixed points of $\sigma$, 
\[
\text{supp}(\sigma) := \{ x \mid x \cdot \sigma \neq x \}
\]
for its support, and 
\[
|\sigma| := |\text{supp}(\sigma)|
\]
for its (unnormalized) Hamming norm. The normalized Hamming norm on $S_n$ is then defined as $|\sigma|_H := \frac{1}{n} |\sigma|$. Also, for $\sigma, \tau \in S_n$ let $d(\sigma, \tau) := |\sigma^{-1} \tau|$ be their Hamming distance and $d_H(\sigma, \tau) := \frac{1}{n} |\sigma^{-1} \tau|$ be their normalized Hamming distance. For $S \subseteq S_n$, write 
\[
\text{diam}(S) := \max_{\sigma, \tau \in S} d(\sigma, \tau)
\]
for the diameter of the set $S$.

We say that $w \in F_r \ast G$ is strong if, when written in the normal form, removing the constants does not lead to any cancellation among the variables—in particular, if $w$ is strong, then either the content is non-trivial or the word was a constant from $G$.

There are some obvious mixed identities for $S_n$. For example, take $\tau$ a transposition, then $[x, \tau]^6$ is such an identity, since $[g, \tau] = \tau^{-8} \tau$ is the product of two transpositions and so either a 3-cycle, the product of two disjoint transpositions, or trivial, where $g \in S_n$ is arbitrary. We notice here that the word with constants $[x, \tau]^6$ has a trivial content. It is one consequence of our main result, Theorem 1, that this must be the case for any mixed identity on $S_n$ of small length. However, there are long words with non-trivial content that are an identity for $S_n$, that is, take $x^{n!}$. We show that a non-trivial word map with constants coming from a short word can have a small image (i.e., with small diameter) only if there are constants involved with small support.

Fix a group $G$ of possible constants. Every word $w \in F_r \ast G$ can be written uniquely in the form 
\[
w = c_0 x_{\varepsilon(1)}^i c_1 \cdots c_{l-1} x_{\varepsilon(l)}^i c_l \in F_r \ast G,
\]
where $x_1, \ldots, x_r$ are the free generators of $F_r$, $\varepsilon(j) = \pm 1, i(j) \in \{1, \ldots, r\} (j = 1, \ldots, l)$, and $c_j \in G (j = 0, \ldots, l)$ are such that there is no cancellation, that is, $x_{\varepsilon(j)}^{i(j)} x_{\varepsilon(j+1)}^{i(j+1)}$ implies that $c_j \neq 1_G$ for $j = 1, \ldots, l - 1$. Then, we set $\ell(w) = l$ and call it the length of $w$. Note that $\ell$ is just the word length where we give elements from $G$ length zero.

Our main result is the following theorem.

**Theorem 1.** Let $w \in F_r \ast S_n$ and consider the associated word map with constants $w : S_n^r \to S_n$. Then, the following holds:

(i) If $w$ has non-trivial content, then 
\[
\frac{\text{diam}(w(S_n^r)) + 1}{n} \geq \frac{1}{2} \exp(-\log(5\ell(w))\ell(w)/2).
\]

(ii) If $w$ is strong and $w \notin S_n$, then 
\[
\frac{\text{diam}(w(S_n^r)) + 1}{n} \geq \frac{1}{2\ell(w)}.
\]

More precisely, if $w \notin S_n$ is arbitrary and the inequality in (ii) is violated, then a constant of size at most $2(\text{diam}(w(S_n^r)) + 1)\ell(w)$ can be found, so that the removal of that constant leads to cancellation of variables.

**Corollary 1.** Assume that $w \in F_r \ast S_n$ ($n \geq 2$) is strong and induces a constant map $S_n^r \to S_n$. Then, either $w \in S_n$ or $\ell(w) \geq \frac{n}{2}$. 


The shortest known strong mixed identities are the laws constructed in [17], they are of length \(\exp(C \log(n)^4 \log \log(n))\). It is an interesting open problem to close this gap.

The second part of Theorem 1 implies the following corollary which is of independent interest.

**Corollary 2.** A metric ultraproduct of symmetric groups \(S_n\) equipped with the normalized Hamming metric, with \(n\) tending to infinity along the chosen ultrafilter, does not satisfy any non-trivial mixed identity.

For completeness, let us briefly recall the notion of the metric ultraproduct of a family of metric groups \((G_n, d_n)_{n \in \mathbb{N}}\) with respect to an ultrafilter \(\mathcal{U}\) on \(\mathbb{N}\). Here, \(d_n\) is a bi-invariant metric on \(G_n\) (\(n \in \mathbb{N}\)). It is defined as the quotient \(\prod_{n \in \mathbb{N}} G_n/N_{\mathcal{U}}\), where \(N_{\mathcal{U}} := \{ (g_n)_{n \in \mathbb{N}} \mid d_n(g_n, 1_{G_n}) \to \mathcal{U} 0\}\) is the normal subgroup of null sequences, and is itself a metric group when equipped with the limit metric.

For more on metric ultraproducts of symmetric groups, see [5, 26, 28]. Note that the previous corollary is in contrast to the usual ultraproduct of symmetric groups, which satisfies a mixed identity \([x, \tau]^6\), where \(\tau\) is an ultraproduct of transpositions as discussed above.

Despite the restrictions on images of word maps with constants given by the length of the words and the structure of the constants involved, we prove in the Appendix that every self-map of a non-abelian simple group is a word map with constants.

## 2 BASIC DEFINITIONS

In this short section, we make some basic definitions, which are needed later. Let \(G\) be a group, the group of possible constants. As above, write the fixed word \(w \in \mathbb{F}_r \ast G\) uniquely in the form

\[ w = c_0 x_1^{\varepsilon(1)} c_1 \cdots c_{l-1} x_1^{\varepsilon(l)} c_l \in \mathbb{F}_r \ast G, \]

where \(\varepsilon(j) = \pm 1\), \(i(j) \in \{1, \ldots, r\}\) (\(j = 1, \ldots, l\)), and \(c_j \in G\) (\(j = 0, \ldots, l\)) are such that there is no cancellation.

We define the sets of indices \(J_0(w), J_+(w), J_-(w) \subseteq \{1, \ldots, l-1\}\) by

\[ J_0(w) := \{ j \mid i(j) \neq i(j + 1)\}, \]

\[ J_+(w) := \{ j \mid i(j) = i(j + 1) \text{ and } \varepsilon(j) = \varepsilon(j + 1)\}, \]

and

\[ J_-(w) := \{ j \mid i(j) = i(j + 1) \text{ and } \varepsilon(j) = -\varepsilon(j + 1)\}. \]

Note that \(J_0(w), J_+(w), J_-(w)\) partition the set \(\{1, \ldots, l-1\}\). We call \(J_-(w)\) the set of critical indices of \(w\) since the removal of a constant \(c_j\) for \(j \in J_-(w)\) leads to cancellation among the variables. The constant \(c_j\) is then called a critical constant. We call \(v\) an elementary reduction of \(w \in \mathbb{F}_r \ast G\), if it is obtained from \(w\) by deleting a critical constant and reducing the outcome.

Now, we focus on word maps with constants in symmetric groups. Keep the notation from above and set \(G := S_n\). For simplicity, we assume that \(c_0 = 1_G = id\). Keep the definition of the sets \(J_-(w), J_+(w)\), and \(J_0(w)\). Define

\[ w_j := x_1^{\varepsilon(1)} c_1 \cdots x_1^{\varepsilon(j)} c_j \]

to be the \(j\)th prefix of \(w\) (\(j = 0, \ldots, l\)). Note that \(w_1 = w\) by construction. For a reduced word \(v \in \mathbb{F}_r \ast S_n\), write \(|v|_i\) for its \(i\)-length, that is, the number of occurrences of \(x_i\) and \(x_i^{-1}\) in its reduced representation (\(i = 1, \ldots, r\)). Let \(w(S_n^\sim) \subseteq S_n\) denote the image of the word map \(w : S_n^\sim \to S_n\) induced by \(w\).
3 PROOF OF THE MAIN RESULT

We are now ready to state the main technical lemma:

Lemma 1. In this setting, assume that \( l > 0 \) and let \( d \geq 1 \) be an integer such that the following hold:

(i) \( n \geq (d - 1)(|w|_{l(j)} + |w|_{l(j+1)}) + |w|_{l(j)} + |w|_{l(j+1)} - 1 \) for all \( j \in J_d(w) \);
(ii) \( n \geq 2(d - 1)(|w|_{l(j)} + |w|_{l(j+1)} - 1) \) for \( j \in J_+(w) \);
(iii) \( |c_j| \geq 2((d - 1)|w|_{l(j)} + |w|_{l(j)} - 1) \) for all \( j \in J_-(w) \);
(iv) \( n \geq d|w|_{l(j)} + 1 \).

Then, we have that \( \text{diam}(w(S'_n)) \geq d \).

To prove this lemma and the results thereafter, we need some additional graph-theoretical terminology: by a directed graph \( G \), we mean an object that consists of a set of vertices \( V(G) \) and a set of edges \( E(G) \) together with maps \( \alpha_{+1}, \alpha_{-1} : E(G) \to V(G) \). For an edge \( e \) of \( G \), we call \( \alpha_{+1}(e) \) its source and \( \alpha_{-1}(e) \) its target; also for \( \varepsilon \in \{ \pm 1 \} \), we call \( \alpha_{\varepsilon}(e) \) the \( \varepsilon \)-source and \( \alpha_{-\varepsilon}(e) \) the \( \varepsilon \)-target of \( e \). Let \( S \) be a set of labels. By an edge-labeling of \( G \) by \( S \) we mean a mapping \( \lambda : E(G) \to S \). If \( G \) is equipped with such a labeling, we call it a directed \( S \)-edge-labeled graph. In this case, an edge \( e \) of \( G \) which is labeled by \( s \in S \) is called an \( s \)-arrow. In this situation, if we exhibit a particular vertex \( v \) of \( G \), we call \((G,v)\) a pointed directed \( S \)-edge-labeled graph.

We will also make use of the following notation for partially defined maps: if \( a : A \to B \) and \( b : B \to C \) are such maps, then their composition \( ab : A \to C \) is the partially defined map given by \( x.ab = (x.a).b \) if \( x.a \) exists and \((x.a).b \) exists, and is undefined otherwise.

Here comes the proof of the previous lemma:

Proof of Lemma 1. We have to prove that, under the assumptions of the lemma, there exist two elements \( \sigma, \tau \in w(S'_n) \) such that \( d(\sigma, \tau) \geq d \). So, let \( \tau \in w(S'_n) \) be an arbitrary element. In the following, we construct \( \sigma \in w(S'_n) \) such that \( d(\sigma, \tau) \geq d \).

Let \( S_n \) act on the \( n \)-element set \( \Omega \) and fix \( d \) arbitrary distinct points \( \omega_1, \ldots, \omega_d \in \Omega \) (this is certainly possible by condition (iv) in the lemma, since \( n \geq d|w|_{l(j)} + 1 > d \) as \( l \geq 1 \)). Now, we inductively construct a family \( (G_k^l, \omega_k^l)_{j,k} \) of pointed directed \( S \)-edge-labeled graphs on the fixed vertex set \( V(G_j^l) := \Omega \) where \( S = \{x_1, \ldots, x_r\} \) is the set of generators of \( F_r \) \( (j = 0, \ldots, l, k = 1, \ldots, d) \). The graphs \( G_k^l \) will be constructed in the following lexicographic order:

\[
G_1^0, G_1^1, G_2^1, \ldots, G_1^l = G_2^0, G_2^1, G_2^2, \ldots, G_2^l = G_3^0, G_3^1, G_3^2, \ldots, G_d^l.
\]

Each such graph \( G_k^l \) will be a partial Schreier graph of \( F_r \) (i.e., a graph that can be completed to a Schreier graph of \( F_r \)) which is obtained from its predecessor by adding a single \( x_{l(j)} \)-arrow \( e_k^l \), starting from the empty graph \( G_0^l \), that is, the graph with \( E(G_0^l) = \emptyset \). Therefore, there will be precisely \( (k - 1)|w||_l + |w||_l \)-many \( x_j \)-arrows in \( G_k^l \) which encode partial injective maps \( \pi_{k,l}^j : \Omega \to \Omega \) \( (l = 1, \ldots, r) \). With this notation, the points \( \omega_k^l \) will be chosen in such a way that

\[
\omega_k^l = \omega_k.w_j(\pi_{k,1}^l, \ldots, \pi_{k,r}^l) \quad \text{and} \quad \omega_k.w_j(\pi_{k,1}^l, \ldots, \pi_{k,r}^l) \neq \omega_k.\tau
\]

for \( j = 0, \ldots, l \) and \( k = 1, \ldots, d \), that is, if \( j \geq 1 \) we have

\[
\alpha_{(j)}(e_k^j) = \omega_k^{j-1} \quad \text{and} \quad \alpha_{(j)}(e_k^j).c_j = \omega_k^j.
\]

Hence, setting \( \pi_1, \ldots, \pi_r \in S_n \) to be extensions of \( \pi_{d,1}^1, \ldots, \pi_{d,r}^1 \), the points \( \omega_k^j \) \( (j = 0, \ldots, l) \) will be precisely the trajectory of \( \omega_k \) under the prefixes of \( w(\pi_1, \ldots, \pi_r) \). Thus, setting \( \sigma := w(\pi_1, \ldots, \pi_r) \), we will have that \( \omega_k.\sigma \neq \omega_k.\tau \) for all \( k = 1, \ldots, d \), so that \( d(\sigma, \tau) \geq d \) as desired.

We are left to carry out the construction of the family \( (G_k^l, \omega_k^l)_{j,k} \) of pointed \( S \)-edge-labeled graphs. Recall that we start with \( G_0^l \) being the empty graph, hence \( E(G_0^l) = \emptyset \) and \( \omega_0^l := \omega_1, \ldots, \omega_d^0 := \omega_d \). Assume that we are to construct \( (G_k^l, \omega_k^l) \) out of the previous data. If \( j = 0 \), there is nothing to do. So assume \( j \geq 1 \) and we are to add the \( x_{l(j)} \)-arrow \( e_k^j \). We assume by induction that we are already given an admissible ‘starting point’ \( \alpha_{(j)}(e_k^j) = \omega_k^{j-1} \) of our edge \( e_k^j \), i.e., \( \omega_k^{j-1} \neq \alpha_{(j)}(e) \) for
any $x_{i(j)}$-arrow $e \in E(G_k^{j-1})$. Our task is to find an admissible ‘end point’ $\alpha_{\varepsilon(j)}(e_k^j)$ of $e_k^j$. This means, we have to ensure that the following conditions are satisfied:

(a) $\alpha_{\varepsilon(j)}(e_k^j) \neq \alpha_{\varepsilon(j)}(e)$ for any $x_{i(j)}$-arrow $e \in E(G_k^{j-1})$;
(b) if $i(j) = i(1)$ and $\varepsilon(j) = -\varepsilon(1)$, then $\alpha_{\varepsilon(j)}(e_k^j) \neq \omega_m$ for all $m = k + 1, \ldots, d$;
(c) if $j < l$, then $\omega_k^j := \alpha_{\varepsilon(j)}(e_k^j) \cdot c_j \neq \alpha_{\varepsilon(j+1)}(e)$ for any $x_{i(j+1)}$-arrow $e \in E(G_k^j) := E(G_k^{j-1}) \cup \{e_k^j\};$
(d) if $j = l$, then $\omega_k^j := \alpha_{\varepsilon(j)}(e_k^j) \cdot c_j \neq \omega_k \cdot \tau$.

Let us briefly explain this: (a) means that $e_k^j$ will not have the same $\varepsilon(j)$-target as any $x_{i(j)}$-arrow in $G_k^{j-1}$, ensuring that $G_k^j$ remains a partial Schreier graph. (b) is necessary, since the $\varepsilon(1)$-sources of the $x_{i(1)}$-arrows $e_m^j$ are already fixed to be $\omega_m$ from the beginning (for all $m = 1, \ldots, d$), so these cannot be used by another $x_{i(1)}$-arrow as its $\varepsilon(1)$-source. (c) is necessary to ensure that $e_k^{j+1}$ will have a valid ‘starting point’ $\omega_k^j = \alpha_{\varepsilon(j+1)}(e_k^{j+1})$ (when $j < l$) which is not already in use as the $\varepsilon(j+1)$-source of another $x_{i(j+1)}$-arrow in $G_k^j$. Finally, (d) ensures that $\omega_k \cdot \sigma \neq \omega_k \cdot \tau$.

Now, we count the number of possibilities to choose the $\varepsilon(j)$-target $\alpha_{\varepsilon(j)}(e_k^j)$ of $e_k^j$ according to (a)–(d): as $G_k^{j-1}$ is a partial Schreier graph, no two of its $x_{i(j)}$-arrows have the same $\varepsilon(j)$-target. Hence, there are precisely

$$n - ((k - 1)|w|_{i(j)} + |w|_{i(j+1)}|l(j)) = n - ((k - 1)|w|_{i(j)} + |w|_{l(j)}| - 1)$$

possible $\varepsilon(j)$-targets for $e_k^j$ that satisfy (a). In the following three cases, we assume that $j < l$, so that (d) is irrelevant.

**Case (i):** $i(j) \neq i(j + 1)$. This means that $j \in J_0(w)$. Then, $G_k^j$ has precisely

$$(k - 1)|w|_{i(j+1)} + |w|_{l(j+1)} = (k - 1)|w|_{i(j+1)} + |w|_{l(j+1)} - 1$$

$x_{i(j+1)}$-arrows, none of which is $e_k^j$, so (c) rules out as many vertices for the $\varepsilon(j)$-target of $e_k^j$. In the worst case, the condition in (b) is satisfied, so that (b) rules out at most $d - k$ more vertices. In total, we get at least

$$n - ((k - 1)|w|_{i(j)} + |w|_{i(j+1)} + |w|_{l(j)} + |w|_{l(j+1)} - 2 + d - k)$$

$$= n - ((k - 1)|w|_{i(j)} + |w|_{i(j+1)} - 1) + |w|_{l(j)} + |w|_{l(j+1)} + d - 3$$

$$\geq n - ((d - 1)|w|_{i(j)} + |w|_{i(j+1)} - 1) + |w|_{l(j)} + |w|_{l(j+1)} + d - 3$$

$$= n - ((d - 1)|w|_{i(j)} + |w|_{i(j+1)} + |w|_{l(j)} + |w|_{l(j+1)} - 2)$$

possible choices for the $\varepsilon(j)$-target of $e_k^j$, which is a positive number by assumption (i) in the lemma.

**Case (ii):** $i(j) = i(j + 1)$ and $\varepsilon(j) = \varepsilon(j + 1)$. This means that $j \in J_4(w)$. Considering (c), there are precisely

$$(k - 1)|w|_{i(j)} + |w|_{l(j)} = (k - 1)|w|_{i(j)} + |w|_{l(j)} - 2$$

$x_{i(j+1)} = x_{i(j)}$-arrows $e$ different from $e_k^j$ in $G_k^j$ (namely the ones in $G_k^{j-1}$) which admit as many vertices as their $\varepsilon(j + 1)$-sources, and the condition $\alpha_{\varepsilon(j)}(e_k^j) \cdot c_j \neq \alpha_{\varepsilon(j)}(e_k^j) \cdot c_j^{-1}$ rules out the vertex $\alpha_{\varepsilon(j)}(e_k^j) \cdot c_j^{-1} = \omega_k^{-1} \cdot c_j^{-1}$. In the worst case, the condition in (b) is satisfied, so that (b) rules out at most $d - k$ further vertices. In total, we get at least

$$n - (2((k - 1)|w|_{i(j)} + |w|_{l(j)} - 2) + 1 + d - k)$$

$$= n - ((k - 1)|w|_{i(j)} - 1) + 2|w|_{l(j)} - 4 + d$$

$$\geq n - ((d - 1)|w|_{i(j)} - 1) + 2|w|_{l(j)} - 4 + d$$

$$= n - (2((d - 1)|w|_{i(j)} + |w|_{l(j)} - 3)$$

possible choices for $\alpha_{\varepsilon(j)}(e_k^j)$, which is positive by assumption (ii) in the lemma.
Case (iii): $i(j) = i(j + 1)$ and $\varepsilon(j) = -\varepsilon(j + 1)$. This means that $j \in J_-(w)$. As in the previous case, (c) splits into two parts: the $x_{i(j+1)} = x_{i(j)}$-arrows different from $e^l_k$ in $G^l_k$ rule out

$$(k - 1)|w|_{i(j)} + |w_{j-1}|_{i(j)} = (k - 1)|w|_{i(j)} + |w_{j+1}|_{i(j)} - 2$$

vertices for the $\varepsilon(j)$-target of $e^l_k$, and the condition $\alpha_{-\varepsilon(j)}(e^l_k), c_j \neq \alpha_{-\varepsilon(j)}(e^l_k)$ rules out the $n - |c_j|$ fixed points of $c_j$. In the worst case, the condition in (b) is satisfied, and (b) rules out $d - k$ further vertices. Hence, we have at least

$$n - (2(k - 1)|w|_{i(j)} + |w_{j+1}|_{i(j)} - 2) + n - |c_j| + d - k$$

$$= |c_j| - ((k - 1)(2|w|_{i(j)} - 1) + 2|w_{j+1}|_{i(j)} - 5 + d)$$

$$\geq |c_j| - ((d - 1)(2|w|_{i(j)} - 1) + 2|w_{j+1}|_{i(j)} - 5 + d)$$

$$= |c_j| - 2((d - 1)|w|_{i(j)} + |w_{j+1}|_{i(j)}) - 4$$

choices for $\alpha_{-\varepsilon(j)}(e^l_k)$, which is positive by assumption (iii) in the lemma.

Case (iv): $j = l$. Then, (c) is irrelevant, (b) can rule out at most $d - k$ vertices for $\alpha_{-\varepsilon(j)}(e^l_k)$ if $(\ast)$ holds, and (d) rules out one further vertex. Hence in total we have at least

$$n - (2(k - 1)|w|_{i(l)} + |w_{l+1}|_{i(l)} - 1 + d - k + 1) \geq n - d|w|_{i(l)}$$

vertices remaining. This is a positive number by assumption (iv) of the lemma. Hence, the proof is complete.

Remark 1. The condition on $\text{diam}(w(S'_{n}))$ is almost sharp: take $g$ to be the $c$-cycle $(1 \cdots c)$ and set $w = g^{x_1}$. Then, $\text{diam}(w(S'_{n})) = \min\{2c, n\}$ and the bound in the lemma gives $\text{diam}(w(S'_{n})) \geq \lfloor c/4 \rfloor$.

We set

$$|w|_{\infty} := \max_{i \in [1, r]} |w|_i \quad \text{and} \quad |w|_{\text{crit}} := \min \left\{ n, \min_{j \in J_-(w)} |c_j| \right\}.$$

The four assumptions in Lemma 1 can obviously by strengthened to the single inequality $|w|_{\text{crit}} \geq 2d|w|_{\infty}$. It follows that

$$\text{diam}(w(S''_{n})) \geq 2|w|_{\text{crit}} - 1$$

for $n \geq 2$.

Note that $|w|_{\infty}$ is obviously bounded by the length $\ell(w)$. We will use the above inequality in the form

$$|w|_{\text{crit}} \leq 2 \cdot (\text{diam}(w(S''_{n})) + 1) \cdot \ell(w).$$

We interpret this inequality as follows: if the diameter of the word image is small, then either there is a small critical constant or the word length is large.

Remark 2. We suppose that a similar lemma can be proved for families like $\text{PSL}_n(p)$ for a fixed prime $p$ and $n$ growing, leading to results comparable to our main theorem. However, we note that for example for the family $\text{PSL}_2(q)$ for $q$ growing, the length of the shortest mixed identity grows linearly in $q$. Note that in this case, the fact that the length of the shortest mixed identity must tend to infinity is already a consequence of the results in [8]. This will be the subject of further work.

Proof of Theorem 1. The bound in (ii) directly follows from the inequality $|w|_{\text{crit}} \leq 2(\text{diam}(w(S''_{n})) + 1)\ell(w)$, since when $w$ is strong we have $|w|_{\text{crit}} = n$ as there is no critical constant. The assertion about the existence of small critical constants is also an immediate consequence of the inequality above.

In order to prove (i) we study the effect of removal of critical constants in more detail. Let $v$ be an elementary reduction of $w \in F_\ast G$, that is, it is obtained from $w$ by deleting a critical constant and reducing the outcome. Note that removing the smallest critical constant $c$ does not change the diameter of the word image by more than $2|c| = 2|w|_{\text{crit}}$, so that given
a small diameter, we can try to iterate elementary reductions. Let \( w \) have non-trivial content and let \( w = w_0, \ldots, w_m \) be a chain of words, where \( w_i \) is an elementary reduction of \( w_{i-1} \) \((i \geq 1)\) by a smallest critical constant, so that there is no such reduction for \( w_m \), that is, \( w_m \) is strong. Note that \( w_i \) no longer denotes the \( i \)th prefix of \( w \). Then, \( \ell(w_i) \leq \ell(w_{i-1}) - 2 \leq \ell(w) - 2i \). We have

\[
|w_{i-1}|_{\text{crit}} \leq 2\ell(w_{i-1})(\text{diam}(w_{i-1}(S_n^r)) + 1)
\]

and so

\[
\text{diam}(w_i(S_n^r)) \leq \text{diam}(w_{i-1}(S_n^r)) + 2|w_{i-1}|_{\text{crit}} \\
\leq \text{diam}(w_{i-1}(S_n^r)) + 4\ell(w_{i-1})(\text{diam}(w_{i-1}(S_n^r)) + 1).
\]

It follows that

\[
\text{diam}(w_i(S_n^r)) + 1 \leq (1 + 4\ell(w_{i-1}))(\text{diam}(w_{i-1}(S_n^r)) + 1).
\]

Hence, we obtain

\[
\text{diam}(w_i(S_n^r)) + 1 \leq (1 + 4\ell(w))(\text{diam}(w(S_n^r)) + 1)
\]

by induction on \( i = 0, \ldots, m \), since \( \ell(w_i) \leq \ell(w) \). By assumption, \( w_m \) is not a constant from \( G = S_n \). Then, by the above inequality

\[
n = |w_m|_{\text{crit}} \leq 2(\text{diam}(w_m(S_n^r)) + 1)\ell(w_m) \\
\leq 2(1 + 4\ell(w))^m(\text{diam}(w(S_n^r)) + 1)\ell(w) \\
\leq 2(1 + 4\ell(w))^{\lfloor \ell(w)/2 \rfloor} \ell(w)(\text{diam}(w(S_n^r)) + 1) \\
\leq 2 \exp(\log(5\ell(w))\ell(w)/2)(\text{diam}(w(S_n^r)) + 1)
\]

as \( m \leq \ell(w)/2 \) and \( \ell(w) \geq 1 \). This is the bound in (i) of Theorem 1.

\[ \square \]

We now turn to metric ultraproducts of symmetric groups and Corollary 2. Assume that \( l \geq 1 \) and let

\[
w_n = x_{i(1)}^{c_{1,n}} \cdots x_{i(l)}^{c_{l,n}} \in F_r * S_n
\]

be a sequence of reduced words of the same structure, but with different constants \( c_{j,n} \) \((j = 1, \ldots, l, n \in \mathbb{N}, n \geq 2)\). Let \( U \) be an ultrafilter on \( \mathbb{N} \) such that \( \text{diam}(w_n(S_n^r))/n \to 0 \) as \( n \to U^* \infty \), that is,

\[
w = x_{i(1)}^{c_{1,n}} \cdots x_{i(l)}^{c_{l,n}} \in U^n \to \infty\]

induces a constant map on the metric ultraproduct of the \( S_n \)’s equipped with the normalized Hamming norm \( |g|_{mH} = |g|/n \). We claim, that then there is an index \( j \in \{1, \ldots, l-1\} \) such that \( i(j) = i(j+1) \) and \( \varepsilon(j) = -\varepsilon(j+1) \), such that \( |c_{j,n}|/n \to U^* 0 \) as \( n \to U^* \infty \). In particular, \( w \) is trivial.

\[ \text{Proof of Corollary 2}. \] By the above \( 2|w|_{\infty}(\text{diam}(w_n(S_n^r))) + 1 \geq |w|_{\infty} \), so that \( |w|_{\infty}/n \to U^* 0 \) as \( n \to U^* \infty \) by the assumption. Hence, by finiteness, there exists an index \( j \in \mathcal{J}(w) \) such that \( |c_{j,n}|/n \to U^* 0 \) as \( n \to U^* \infty \). In particular, starting with a potential mixed identity in the metric ultraproduct, the previous argument implies that a critical constant must be trivial. This is a contradiction and the proof is complete.

\[ \square \]

**Acknowledgments**

We thank Robert Kaak for his friendship. Moreover, we thank Christoph Schulze and Vadim Alekseev for discussions on the topic. We thank Ben Steinberg and Anton Klyachko for comments on a first draft of this paper. Finally, we thank the unknown referee for valuable comments. This research was supported by the ERC Consolidator Grant No. 681207.
REFERENCES

[1] Z. Arad, J. Stavi, and M. Herzog, Powers and products of conjugacy classes in groups, Lecture Notes in Mathematics, vol. 1112, Springer, Berlin, 1985, pp. 6–51.

[2] N. Avni, T. Gelander, M. Kassabov, and A. Shalev, Word values in \( p \)-adic and adelic groups, Bull. London Math. Soc. 45 (2013), no. 6, 1323–1330.

[3] T. Bandman, S. Garion, and F. Grunewald, On the surjectivity of Engel words on \( \text{PSL}(2, q) \), Groups Geom. Dyn. 6 (2012), no. 3, 409–439.

[4] H. Bradford and A. Thom, Short laws for finite groups and residual finiteness growth, Trans. Am. Math. Soc. 371 (2019), no. 9, 6447–6462.

[5] G. Elek and E. Szabó, Sofic groups and direct finiteness, J. Algebra 280 (2004), no. 2, 426–434.

[6] A. Elkasapy and A. Thom, About Goto’s method showing surjectivity of word maps, Indiana Univ. Math. J. 63 (2014), no. 5, 1553–1565.

[7] S. Garion and A. Shalev, Commutator maps, measure preservation, and \( T \)-systems, Trans. Am. Math. Soc. 361 (2009), no. 9, 4631–4651.

[8] N. Gordeev, B. Kunyavskii, and E. Plotkin, Word maps and word maps with constants of simple algebraic groups, Dokl. Math. 94 (2016), no. 3, 632–634.

[9] N. Gordeev, B. Kunyavskii, and E. Plotkin, Word maps on perfect algebraic groups, Int. J. Algebr. Comput. 28 (2018), no. 8, 1487–1515.

[10] N. L. Gordeev, Freedom in conjugacy classes of simple algebraic groups and identities with constants, Algebr. Anal. 9 (1997), no. 4, 63–78.

[11] R. Guralnick and P. H. Tiep, Effective results on the Waring problem for finite simple groups, Am. J. Math. 137 (2015), no. 5, 1401–1430.

[12] R. Guralnick, M. W. Liebeck, E. A. O’Brien, A. Shalev, P. H. Tiep, Surjective word maps and Burnside’s \( p^aq^b \) theorem, Invent. Math. 213 (2018), no. 2, 589–695.

[13] G. Horváth and C. L. Nehaniv, Length of polynomials over finite groups, J. Comput. Syst. Sci. 81 (2015), no. 8, 1614–1622.

[14] C. Y. Hui, M. Larsen, and A. Shalev, The Waring problem for Lie groups and Chevalley groups, Isr. J. Math. 210 (2015), no. 1, 81–100.

[15] M. Istinger and H. K. Kaiser, A characterization of polynomially complete algebras, J. Algebra 56 (1979), no. 1, 103–110.

[16] A. Klyachko and A. Thom, New topological methods to solve equations over groups, Algebraic Geom. Topol. 17 (2017), no. 1, 331–353.

[17] G. Kozma and A. Thom, Divisibility and laws in finite simple groups, Math. Ann. 364 (2016), no. 1-2, 79–95.

[18] M. Larsen, Word maps have large image, Isr. J. Math. 139 (2004), no. 1, 149–156.

[19] M. Larsen and A. Shalev, Word maps and Waring type problems, J. Am. Math. Soc. 22 (2009), no. 2, 437–466.

[20] M. Larsen and A. Shalev, On the distribution of values of certain word maps, Trans. Am. Math. Soc. 368 (2016), no. 3, 1647–1661.

[21] M. Larsen and P. H. Tiep, Words, Hausdorff dimension and randomly free groups, Math. Ann. 371 (2018), 1409–1427.

[22] M. Larsen, A. Shalev, and P. H. Tiep, Waring problem for finite quasi-simple groups, Int. Math. Res. Not. 2013 (2012), no. 10, 2323–2348.

[23] M. W. Liebeck and A. Shalev, Diameters of finite simple groups: sharp bounds and applications, Ann. Math. (2001), 383–406.

[24] A. Lubotzky, Images of word maps in finite simple groups, Glasg. Math. J. 56 (2014), no. 2, 465–469.

[25] W. D. Maurer and J. L. Rhodes, A property of finite simple non-abelian groups, Proc. Am. Math. Soc. 16 (1965), 552–554.

[26] N. Nikolov, J. Schneider, and A. Thom, Some remarks on finitarily approximable groups, J. Éc. Polytech. Math. 5 (2018), 239–258.

[27] M. Nitsche and A. Thom, Universal solvability of group equations, J. Group Theory 25 (2022), no. 1, 1–10.

[28] V. Pestov, Hyperlinear and sofic groups: a brief guide, Bull. Symb. Log. 14 (2008), no. 4, 449–480.

[29] J. Schneider, On ultraproducts of compact quasi-simple groups, PhD thesis, TU Dresden, 2019.

[30] A. Thom, Convergent sequences in discrete groups, Can. Math. Bull. 56 (2013), no. 2, 424–433.

[31] A. Thom, About the length of laws for finite groups, Isr. J. Math. 219 (2017), no. 1, 469–478.

[32] G. M. Tomanov, Generalized group identities in linear groups, Math. USSR-Sb. 51 (1985), no. 1, 33.

How to cite this article: J. Schneider and A. Thom, Word maps with constants on symmetric groups, Math. Nachr. 297 (2024), 165–173. https://doi.org/10.1002/mana.202300152

APPENDIX

It is well known that every map \( F_q \rightarrow F_q \) is realized by a polynomial from \( F_q[\mathbb{X}] \) of degree less than \( q \)—in fact this is a characterization of fields among finite rings, see [15]. We prove the analogous result for non-abelian simple groups. We denote by the covering number \( cn(G) \) of \( G \) the minimal \( m \), such that \( C^m = G \) for all non-trivial conjugacy classes \( C \). Various bounds on \( cn(G) \) for finite simple groups \( G \) can be found in the literature, see, for example, the seminal work of Liebeck–Shalev [23], and references therein.

Theorem A1. A non-abelian finite group is simple if and only if every map \( G \rightarrow G \) is a word map with constants. Every such map can be represented by a word of length at most \( O(|G|^3 \, cn(G)) \).

Remark A1. An abelian group \( G \) has the above property (i.e., each map \( G \rightarrow G \) comes from a word with constants) if and only if \( |G| \leq 2 \).
Remark A2. In the theorem, the term $\text{cn}(G)$ can also be replaced by the covering diameter $\text{cd}(G)$ of $G$, which is the minimal number $m$ such that $(\{1_G\} \cup C \cup C^{-1})^m = G$ for all non-trivial conjugacy classes $C$ of $G$. Clearly, $\text{cd}(G) \leq \text{cn}(G)$.

Proof. The evaluation map $\text{ev} : F_1 \ast G = \langle x \rangle \ast G \to G^G$ is a homomorphism. Also, if $\pi_{[g,h]} : G^G \to G^{[g,h]}$ is the projection onto $g$ and $h$, then $\pi_{[g,h]} \circ \text{ev}$ is surjective: indeed, by Goursat’s lemma and since $G$ is assumed to be simple, it is enough to show that its image is not the graph of an automorphism. However, this cannot be the case, since the constants $G$ map to the diagonal subgroup, and $x$ maps to a non-diagonal subgroup. This implies that $\text{ev}$ must be surjective too.

By using iterated commutators, one can give a more explicit construction: we prove that for any subset $S \subseteq G$ of size $m \leq 2^e$ and an element $g \in G \setminus S$, there is a word map $w_{g,S} \in F_1 \ast G$ of length $4^e$ such that the induced map $w_{g,S} : G \to G$ satisfies $w_{g,S}(s) = 1_G$ for all $s \in S$ and $w_{g,S}(g) \neq 1_G$.

Indeed, this is true when $e = 0$ and $S = \{s\}$: simply take $w_{g,[s]} = xs^{-1}$. For $e \geq 1$, write $S = S_1 \cup S_2$ as a union of the subsets $S_1$ and $S_2$ such that $|S_1|, |S_2| \leq 2^{e-1}$. Then, set $w_{g,S} := [w_{g,S_1}^a, w_{g,S_2}^b]$ where $a, b \in G$ are chosen such that $[w_{g,S_1}(g)^a, w_{g,S_2}(g)^b] \neq 1_G$. This is possible since $(w_{g,S_2}(g))^G$ generates the whole $G$ by simplicity as $w_{g,S_2}(g) \neq 1_G$, whereas $C_G(w_{g,S_1}(g))$ is a proper subgroup. Hence,

$$w_{g,S}^G(g) \not\subseteq C_G(w_{g,S_1}(g)).$$

We have that $\ell(w_{g,S}) \leq 2(\ell(w_{g,S_1}) + \ell(w_{g,S_2})) \leq 4^e$.

Then setting $e := \lceil \log_2(|G| - 1) \rceil$, one can take $S = G \setminus \{g\}$, so that $w_{g,S}(s) = 1_G$ for $s \in S$ and $w_{g,S}(g) \neq 1_G$ as desired. Now, we can multiply together conjugates of the map $w_{g,S}$ to get a map $\delta_{g,h} \in F_1 \ast G$ such that $\delta_{g,h}(s) = 1_G$ for $s \in S = G \setminus \{g\}$ and $\delta_{g,h}(h) = h$ for a chosen $h \in G$. Then $w = \prod_{g \in G} \delta_{g,f(g)} = f$ as maps $w, f : G \to G$ where $f$ is arbitrary. So, $f$ is a word map. Here

$$\ell(w) \leq |G||\log_2(|G| - 1)| \text{cn}(G) \leq 4|G||(|G| - 1)^2 \text{cn}(G) < 4|G|^3 \text{cn}(G).$$

Conversely, if any map $f : G \to G$ is a word map, then for $g \neq 1_G$, the map $\delta_{g,h} : G \to G$ given by

$$\delta_{g,h}(x) = \begin{cases} 1_G & \text{if } x \neq g \\ h & \text{if } x = g \end{cases}$$

is induced by some word $w \in F_1 \ast G$. Since $w(1_G) = \delta_{g,h}(1_G) = 1_G$, we have that $w(g) = \delta_{g,h}(g) = h \in \langle g \rangle$. But since $h$ was arbitrary, we have $\langle g \rangle = G$, so that $G$ is simple, since $g \neq 1_G$ was arbitrary as well. \[\square\]

By [1], Chapter 3, we have $\text{cn}(A_n) = \lceil n/2 \rceil$ for $n \geq 6$. Thus, we get the following estimate for alternating groups.

Corollary A1. For $n \geq 5$, every map $A_n \to A_n$ can be represented by a word map with constants of length $O((n!)^3 n)$.

Remark A3. There are at most $3^{|G|^{[G]} + 1}$ words with constants of length at most $l$ in $F_1 \ast G = \langle x \rangle \ast G$. So in order that all self-maps $G \to G$ can be represented as word maps coming from words of length at most $l$, we must have

$$|G|^{[G]} \leq 3^{|G|^{[G]} + 1}.$$  

This implies a linear lower bound on the minimal length of words with constants that can represent all possible maps.

Remark A4. It was pointed out to us by Ben Steinberg and Anton Klyachko, that the qualitative aspect of Theorem A1 is a well-known result [25]. Similar but weaker bounds have been obtained in [13]. Using methods similar to the ones in the proof of Theorem A1 one can show that every map $G' \to G'$ arises as a word map with constants of length $O(|G'|^{r+2} \text{cn}(G'))$. Indeed, one can use the maps of length at most $O(|G'|^2)$ to get maps that see just one coordinate of $G'$. Taking $[\log_2(r)]$ commutators of those maps, one gets a map that sees just one element of $G'$. This then has length $O(r^2 |G'|^2)$. Hence, we get a bound

$$O(|G'|^{r+2} r^2 \text{cn}(G)).$$

This is exactly the bound in [13] for $A_n$, but the authors do not write it out for other non-abelian simple groups.