

SOLVABILITY OF SYMMETRIC WORD EQUATIONS IN POSITIVE DEFINITE LETTERS

SCOTT N. ARMSTRONG AND CHRISTOPHER J. HILLAR

ABSTRACT

Let $S(X, B)$ be a symmetric ('palindromic') word in two letters X and B . A theorem due to Hillar and Johnson states that for each pair of positive definite matrices B and P , there is a positive definite solution X to the word equation $S(X, B) = P$. They also conjectured that these solutions are finite and unique. In this paper, we resolve a modified version of this conjecture by showing that the Brouwer degree of such an equation is equal to 1 (in the case of real matrices). It follows that, generically, the number of solutions is odd (and thus finite) in the real case. Our approach allows us to address the more subtle question of uniqueness by exhibiting equations with multiple real solutions, as well as providing a second proof of the result of Hillar and Johnson in the real case.

1. Introduction

In this paper, we consider a natural matrix generalization to the elementary scalar equation

$$bx^s = p,$$

in which $b > 0$, $p \geq 0$, $s \in \mathbb{Z}_+$ and x is a nonnegative real indeterminate. One difficulty with an extension is dealing with matrix noncommutativity, whereas another is determining what should be meant by the words 'real' and 'nonnegative'. Fortunately for us, the latter concerns have already long been addressed: the natural matrix interpretation of the reals comprises the Hermitian matrices, whereas nonnegative numbers correspond to those complex Hermitian matrices with all nonnegative eigenvalues, the so-called *positive semidefinite* matrices, and positive numbers correspond to such matrices with all positive eigenvalues, the so-called *positive definite* matrices. The issue of noncommutativity, however, is of a more subtle nature, and we first introduce some notation before addressing it.

Fix a positive integer k , and let $W = W(X, B_1, \dots, B_k)$ be a word in the letters X and B_1, \dots, B_k . The *reversal* of W is the word written in reverse order, and it is denoted by W^* . A word is *symmetric* if it is identical to its reversal (in other contexts, the term 'palindromic' is also used). As we shall soon see in Sections 2 and 3, formulating our generalization requires restriction to a special class of words. For the purposes of this work, an *interlaced word* $W = W(X, B_1, \dots, B_k)$ in the *interlacing letter* X is a juxtaposition of powers of letters that alternate in powers of X . More precisely, an interlaced word is an expression of the form

$$W = B_{i_1}^{q_1} \prod_{j=1}^m X^{p_j} B_{i_{j+1}}^{q_{j+1}}, \quad (1)$$

in which the exponents $p_j > 0$, $q_j \geq 0$ are nonnegative integers, $m \geq 1$, and $\{i_1, \dots, i_{m+1}\} \subseteq \{1, \dots, k\}$. (Here, of course, we consider the 0th power of a letter to be the empty word, the identity element of the monoid.) For example, the word $B_1 X B_3^2 X^2 B_2^3 X^5$ is interlaced, whereas the word $X B_1 B_2 X B_2 B_1 X$ is not. The integer $s = p_1 + \dots + p_m$ is called the *degree* of the interlaced word W .

Received 27 July 2006; revised 16 March 2007; published online 20 November 2007.

2000 *Mathematics Subject Classification* 15A24, 15A57; 15A18, 15A90.

The work of the second author is supported under a National Science Foundation Postdoctoral Fellowship.

The interlacing letter X is distinguished and is to be viewed as an indeterminate $n \times n$ positive semidefinite matrix, whereas the letters B_1, \dots, B_k correspond to fixed $n \times n$ positive definite matrices. For convenience, the letters X and B_i will also represent the substituted matrices (the context will make the distinction clear). When $k = 1$, the set of interlaced words is simply the set of all words in two letters containing at least one X . For notational simplicity, when k is understood, we write $W(X, B_i)$ in place of $W(X, B_1, \dots, B_k)$.

Returning to our example, notice that there is a unique nonnegative solution to the equation $bx^s = p$ for every pair of positive b and nonnegative p ; we would like to generalize this observation. Our introductory remarks help us to make the following definition.

DEFINITION 1.1. A *symmetric word equation* is an equation, $S(X, B_i) = P$, in which $S(X, B_i)$ is an interlaced symmetric word. If the B_i are positive definite and P is positive semidefinite, then any positive semidefinite matrix X for which the equation holds is called a *solution* to the symmetric word equation.

A symmetric word equation will be called *solvable* if there exists a solution for every positive definite $n \times n$ matrix B_i and $n \times n$ positive semidefinite P . Moreover, if each such B_i and P gives rise to a unique solution, then the equation will be called *uniquely solvable*. We are motivated by the following striking result.

THEOREM 1.2 (Hillar and Johnson). *Every symmetric word equation is solvable. Moreover, if the parameters P and B_i are real, then there is a real solution.*

Theorem 1.2 first appeared in [13, Theorem 7.1] with an argument that involved fixed-point methods. The authors of [13] also conjectured that symmetric word equations have unique solutions.

CONJECTURE 1.3. Every symmetric word equation is uniquely solvable.

There is much evidence to support Conjecture 1.3. For instance, there are large classes of equations that are uniquely solvable (see Section 5, where we encounter the class of *totally symmetric word equations*), and recently, Lawson and Lim [21] have verified the conjecture in the case that the degree of $S(X, B_i)$ is not greater than 5. Their approach utilizes the Riemannian metric on the set of positive definite matrices and Banach's fixed-point theorem. In addition, every numerical investigation that we are aware of has failed to produce multiple solutions (see also Remark 11.3).

Symmetric word equations are not only natural from a theoretical perspective; they also arise in many other contexts. For instance, they play a role in recent attacks [11, 12, 14] on the Bessis–Moussa–Villani trace conjecture [3], a long-standing problem in statistical physics. A brief overview of this application is given in Section 4.

One well-known matrix equation is the Riccati equation:

$$XBX^T = P. \tag{2}$$

In general, the Riccati equation (2) has many solutions; however, when B and P are positive definite and we seek positive definite solutions X , it is equivalent to the symmetric word equation $XBX = P$. The unique positive definite solution X to this equation is given by

$$X = P^{1/2}(P^{-1/2}B^{-1}P^{-1/2})^{1/2}P^{1/2}. \tag{3}$$

This fact has been observed by many authors independently (see, for example, [4, 8, 12, 19, 26, 30]). The right-hand side of (3) is called the *geometric mean* of P and B^{-1} and is written as $P\#B^{-1}$. The Riccati equation and the corresponding geometric mean are

ubiquitous. They appear, for example, in work on matrix inequalities [1, 2], the theory of optimal transportation [8, 19, 26, 27, 30], convex optimization and control theory [10, 29], and the geometry of non-compact symmetric spaces [20, 23].

In this paper, we resolve Conjecture 1.3 negatively in the case $n \geq 3$. The conjecture remains open for 2×2 matrices (although see Theorem 11.4 for a proof of uniqueness in a special case).

THEOREM 1.4. *There are symmetric word equations of degree 6 which have multiple real 3×3 positive definite solutions.*

Theorem 1.4 shows that the result of Lawson and Lim [21] is optimal. Although uniqueness fails in general, our approach allows us to verify that these equations are still very well-behaved in the following sense.

THEOREM 1.5. *Fix an interlaced symmetric word S , positive definite matrices B_1, \dots, B_k , and a positive semidefinite P . Then solutions X to the word equation $f(X) = S(X, B_i) = P$ are bounded. In addition, if we restrict to real matrices and P is invertible, then there is a bounded open subset U of real positive definite matrices (containing all real solutions) such that*

$$\deg(f, U, P) = 1.$$

Here, $\deg(f, U, P)$ is the Brouwer degree of f at P with respect to U ; in a vague sense, it gives a topological measure of the number of solutions inside U to the equation $f(X) = P$. It is in this sense that Theorem 1.5 verifies the intuition of Conjecture 1.3. Theorem 1.5 is powerful enough to show that while uniqueness fails in general, generically, the number of solutions to a symmetric word equation is finite.

COROLLARY 1.6. *Fix real positive definite matrices B_1, \dots, B_k . Then, for almost every real positive definite matrix P , the symmetric word equation*

$$S(X, B_i) = P$$

has an odd (and thus finite) number of real positive definite solutions X .

Proof. By Theorems 1.5 and 7.3, at any regular value P of the map $X \mapsto S(X, B_i)$, the equation $S(X, B_i) = P$ has an odd number of solutions X . By Sard's theorem, the set of regular values is a set of full measure, completing the proof. \square

The hypothesis of Theorem 1.5 requiring the symmetric word to be interlaced cannot be dropped: there exist non-interlaced symmetric word equations with an unbounded set of solutions (see Example 2.5). Theorem 1.5 also implies a special case of Theorem 1.2, giving a second proof of existence in the real case.

COROLLARY 1.7. *Every real symmetric word equation has a real positive semidefinite solution.*

Proof. The result follows from Theorems 1.5 and 6.2 and Lemma 7.1. \square

We remark that if the symmetric word is not interlaced, a solution might not exist (see Example 2.6).

The proof of Theorem 1.5 is the content of Sections 8, 9, and 10. The arguments in the proof often employ the reductions found in Section 6. Some consequences of Theorem 1.5 are explored in Section 11, including a proof of Theorem 1.4. In Sections 2 and 3, we explain our

restriction to interlaced symmetric words, and Sections 4 and 5 are devoted to applications and a special class of uniquely solvable words, respectively. In Section 7, we review the theory of Brouwer degree.

2. Examples

The simplest instance of a symmetric word equation arises in the following example [17, p. 405]; it is the most straightforward generalization of the scalar case.

EXAMPLE 2.1. Let P be any positive semidefinite matrix and let $S(X)$ be the word X^m , for a positive integer m . Then, there is a unique positive semidefinite solution to the equation $S(X) = P$. In fact, writing $P = UDU^*$ for a unitary matrix U and a nonnegative diagonal matrix D , we have $X = UD^{1/m}U^*$.

Our next example is the Riccati equation, which we encountered in the introduction.

EXAMPLE 2.2. Given positive definite B and positive semidefinite P , the equation $XBX = P$ has a unique positive semidefinite solution X , given by

$$X = B^{-1}\#P = B^{-1/2}(B^{1/2}PB^{1/2})^{1/2}B^{-1/2}.$$

Uniqueness can be deduced from the proof of Proposition 5.2, in which a large class of word equations are shown to be uniquely solvable. When P is invertible, this solution can also be expressed as

$$X = P\#B^{-1} = P^{1/2}(P^{-1/2}B^{-1}P^{-1/2})^{1/2}P^{1/2};$$

that is, the geometric mean satisfies $A\#B = B\#A$ for all positive definite matrices A and B . At first glance, this is surprising, since the expression

$$A\#B = A^{1/2}(A^{-1/2}BA^{-1/2})^{1/2}A^{1/2}$$

does not appear to be symmetric in A and B .

As promised, we now explain why we restrict our attention to interlaced symmetric words. A first obstacle in generalizing the scalar case is that most words do not evaluate to positive semidefinite matrices upon substitution. One simple example is the word XB , which does not even have to be Hermitian when X and B are positive definite. Similarly, the unique matrix solution X of the equation $XB = P$ is not, in general, positive semidefinite. It turns out that the right class of words to consider comprises the symmetric ones, and this is evidenced by the following discussion.

Recall that two $n \times n$ matrices X and Y are said to be *congruent* if there is an invertible $n \times n$ matrix Z such that $Y = Z^*XZ$ (here, C^* denotes the *conjugate transpose* of a complex matrix C); and that congruence on Hermitian matrices preserves inertia (the ordered triple consisting of the number of positive, negative, and zero eigenvalues) and, thus, positive definiteness [17, p. 223]. A symmetric word evaluated at positive definite matrices is inductively congruent to the ‘center’, positive definite matrix. We conclude that the following lemma holds.

LEMMA 2.3. *A symmetric word evaluated at positive definite matrices is positive definite.*

A more careful examination (or a simple continuity argument) also proves the following statement.

LEMMA 2.4. *A symmetric word evaluated at positive semidefinite matrices is positive semidefinite.*

Conversely, it may be shown that symmetric words are the only words that are positive definite for all positive definite substitutions (see Section 3 for a proof). In light of these facts, restricting our consideration to symmetric words seems appropriate.

Next, we discuss the difficulties that arise when considering non-interlaced symmetric words. As the following examples demonstrate, both finiteness and existence may fail even when $k = n = 2$ and $s = 3$.

EXAMPLE 2.5. Let $S(X, B_1, B_2) = XB_1B_2XB_2B_1X$ and set

$$B_1 = \begin{bmatrix} 3 & -1 \\ -1 & 1 \end{bmatrix}, \quad B_2 = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{and} \quad P = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

Then, as is easily verified, the equation $S(X, B_1, B_2) = P$ has symmetric solutions

$$X = \begin{bmatrix} 0 & 0 \\ 0 & x \end{bmatrix} \quad \text{and} \quad X = \begin{bmatrix} x/5 & -x \\ -x & 5x \end{bmatrix},$$

in which x is an arbitrary real number. In particular, there are infinitely many positive semidefinite solutions (in two distinct unbounded solution classes). Notice also that the kernel of a solution X and that of P can be different. For interlaced words, this situation cannot occur (see Lemma 6.1).

EXAMPLE 2.6. Let S and B_1, B_2 be as in the previous example, but instead set

$$P = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then, there are no positive semidefinite solutions to $S(X, B_1, B_2) = P$. To verify this, suppose that

$$X = \begin{bmatrix} e & f \\ g & h \end{bmatrix}$$

is a complex solution to $S(X, B_1, B_2) = P$. Computing the ideal generated by the four consequent polynomial equations (using MAPLE or MACAULAY 2 to find the reduced Gröbner basis), we find that it is the entire ring $\mathbb{C}[e, f, g, h]$. In particular, there are no matrix solutions over \mathbb{C} to the given equation, much less positive semidefinite ones.

3. Relations between positive definite words

In this section, we explain our restriction to symmetric words. Specifically, we prove that a word $W(A, B)$ in two letters A and B is positive definite for all positive definite substitutions if and only if the word is symmetric.

We begin by illustrating some of the subtlety of the problem. Let B and P be positive definite matrices. In Example 2.2, we saw that

$$P^{1/2}(P^{-1/2}B^{-1}P^{-1/2})^{1/2}P^{1/2} = B^{-1/2}(B^{1/2}PB^{1/2})^{1/2}B^{-1/2},$$

even though both expressions are quite different. In fact, both sides of the above equality are the unique solution X to the symmetric word equation,

$$S(X, B) = XBX = P.$$

Fortunately, such behavior does not occur with words, as the following discussion illustrates.

Let \mathcal{W} be the set of words in two letters A and B , and fix a, b to be two $n \times n$ complex matrices. Consider the *evaluation homomorphism* $\text{Eval}_{a,b} : \mathcal{W} \rightarrow \mathbb{M}_n(\mathbb{C})$ which sends a word $W(A, B)$ to the matrix $W(a, b)$ produced by substituting the matrices a and b for the letters A and B , respectively. By convention, the empty word is sent to the identity matrix by this map. We describe a pair of positive definite a and b for which this function is injective.

LEMMA 3.1. *The map $\text{Eval}_{a,b}$ is injective when*

$$a = \begin{bmatrix} 3 & 1 \\ 1 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}.$$

Proof. Let a, b be the matrices in the statement of the lemma, and let W_1 and W_2 be two words for which $W_1(a, b) = W_2(a, b)$; we must show that W_1 and W_2 are the same word. If either W_1 or W_2 is the empty word, then the claim is clear (take a determinant). Furthermore, since a and b are invertible, we may suppose that $W_1 = AU$ and $W_2 = BV$ for some words U and V .

Let x and y be indeterminates. Given a word W , we set

$$\begin{bmatrix} W^x x + W^y y \\ W_x x + W_y y \end{bmatrix} = W(a, b) \begin{bmatrix} x \\ y \end{bmatrix},$$

for natural numbers $W^x, W^y, W_x,$ and W_y . Notice that by our choice of a and b , we cannot have both W^x and W^y equal to zero. A direct computation shows that $(AU)^x - (AU)_x = 2U^x$ and that $(BV)^x - (BV)_x = -2V_x$. By assumption, these two numbers are equal so that $U^x + V_x = 0$. Since these two quantities are nonnegative integers, it follows that $U^x = V_x = 0$. Similarly, the equality $(AU)^y - (AU)_y = (BV)^y - (BV)_y$ implies that $U^y = V_y = 0$. This contradiction finishes the proof. □

COROLLARY 3.2. *The following are equivalent for a word W .*

- (1) *W is positive definite for all substitutions of positive definite A and B .*
- (2) *W is Hermitian for all substitutions of positive definite A and B .*
- (3) *W is Hermitian for all 2×2 substitutions of positive definite A and B .*
- (4) *W is symmetric ('palindromic').*

In particular, if a word is Hermitian for all 2×2 substitutions of positive definite A and B , then the word is necessarily positive definite for all such substitutions.

Proof. (1) \Rightarrow (2) \Rightarrow (3) is clear. If $W(A, B)$ is always Hermitian for 2×2 positive definite A and B , then $W(A, B)^* = W(A, B)$ for all such A and B . However, Lemma 3.1 shows that W^* and W must be identical as words. It follows that W is symmetric. This proves that (3) \Rightarrow (4). Finally, if W is symmetric, then Lemma 2.3 implies that W will always be positive definite for any positive definite A and B . This completes the proof. □

4. An application

We first encountered symmetric word equations when studying a trace conjecture [12] involving words in two letters A and B (see also [15, 16]).

CONJECTURE 4.1. *A word in two letters A and B has positive trace for every pair of real positive definite A and B if and only if the word is symmetric or a product (juxtaposition) of two symmetric words.*

For each solvable symmetric word equation, one can identify an infinite class of words that admit real positive definite matrices A and B giving those words a negative trace. The following is a brief description of this application. Consider the word $W = BABAAB$, which is neither symmetric nor a product of two symmetric words. In light of Conjecture 4.1, we would like to verify that there exist real positive definite matrices A and B giving W a negative trace. This is surprisingly difficult, as the methods in [12] show. Resulting A and B that exhibit a

negative trace are, for example,

$$A_1 = \begin{bmatrix} 1 & 20 & 210 \\ 20 & 402 & 4240 \\ 210 & 4240 & 44\,903 \end{bmatrix} \quad \text{and} \quad B_1 = \begin{bmatrix} 36\,501 & -3820 & 190 \\ -3820 & 401 & -20 \\ 190 & -20 & 1 \end{bmatrix}.$$

We will run into these matrices again in our proof of Theorem 1.4.

Consider now the following extension. Let T be the word given by $T = S_1S_2$, in which S_1 and S_2 are symmetric words in the letters A and B . If the simultaneous word equations

$$\begin{aligned} S_1(A, B) &= B_1, \\ S_2(A, B) &= A_1 \end{aligned}$$

may be solved for positive definite A and B given positive definite A_1 and B_1 , then the word TTT^* can have negative trace. Specializing to the case that S_2 is the word A , we have the following corollary.

COROLLARY 4.2. *Let $S = S(A, B)$ be any symmetric word with at least one B . Then the word $SASAAS$ admits real positive definite matrices A and B giving it negative trace.*

Proof. The matrix $B_1A_1B_1A_1B_1$ has negative trace. Using Corollary 1.7, the equation $S(A_1, X) = B_1$ has a real positive definite solution $X = B_2$. The two matrices $B = B_2$ and $A = A_1$ are then the desired witnesses. □

Conjecture 4.1, while interesting in its own right, arises from an old problem in statistical physics, the Bessis–Moussa–Villani conjecture. In [3], while studying partition functions of quantum mechanical systems, those authors postulated a conjecture regarding a positivity property of traces of matrices. If this property holds, then explicit error bounds in a sequence of Padé approximants follow. Recently, in [22], and as previously communicated to the authors of [12], the conjecture of [3] was reformulated by Lieb and Seiringer as a question about the traces of certain sums of words in two positive definite matrices.

CONJECTURE 4.3 (Bessis–Moussa–Villani). The polynomial $p(t) = \text{Tr}[(A + tB)^m]$ has all positive coefficients whenever A and B are $n \times n$ positive definite matrices.

The coefficient of t^k in $p(t)$ is the trace of $H_{m,k}(A, B)$, the sum of all words of length m in A and B , in which k instances of B appear. Since its introduction in [3], many partial results and substantial computational experimentation have been given [5, 6, 12, 14, 28], all in favor of the conjecture’s validity. However, despite much work, very little is known about the problem, and it has remained unresolved except in very special cases. Until recently, even the case $m = 6$ and $n = 3$ was unknown. In this case, all coefficients, except $\text{Tr}[H_{6,3}(A, B)]$ were known to be positive [12]. The remaining coefficient $\text{Tr}[H_{6,3}(A, B)]$ can be shown to be positive, but the proof requires notably different methods [14]. The difficulty is that some summands of $H_{6,3}(A, B)$ can have negative trace, precisely the types of words such as $BABAAB$ considered above. The matrices A_1 and B_1 above were also used by Hansen [9] in his approach to this trace conjecture.

A recent advance [11] has been the derivation of a pair of equations satisfied by A and B with Euclidean norm 1 that minimize a coefficient $\text{Tr}[H_{m,k}(A, B)]$:

$$\begin{aligned} AH_{m-1,k}(A, B) &= A^2\text{Tr}[AH_{m-1,k}(A, B)]; \\ BH_{m-1,k-1}(A, B) &= B^2\text{Tr}[BH_{m-1,k-1}(A, B)]. \end{aligned}$$

It is possible that some of the techniques developed here can be applied to these more general types of word equations.

5. A class of uniquely solvable equations

In this section, we describe a class of words that are uniquely solvable with solutions that can be constructed explicitly. These words generalize those found in Examples 2.1 and 2.2.

DEFINITION 5.1. A symmetric word is called *totally symmetric* if it can be expressed as a composition of maps of the form:

- (1) $\pi_{m,B_i}(W) = (WB_i)^m W$, m a positive integer,
- (2) $\varphi_m(W) = W^m$, m a positive integer,
- (3) $\mathcal{C}_{B_i}(W) = B_i W B_i$,

applied to the letter X .

For example, the word $W = B_1 X^2 B_2 X^2 B_2 X^2 B_1$ may be expressed as the composition, $\mathcal{C}_{B_1} \circ \pi_{2,B_2} \circ \varphi_2(X)$. The utility of this definition becomes clear from the following proposition.

PROPOSITION 5.2. For every totally symmetric word $S(X, B_i)$ and every positive definite B_i and positive semidefinite P , the equation $S(X, B_i) = P$ has a unique positive semidefinite solution X .

Proof. We induct on the number of compositions involved in the word S , the base case $S = X$ being trivial. If $S = \varphi_m(W)$ for some word W , then $W = P^{1/m}$ is a smaller totally symmetric word equation and any solution X to $S(X, B_i) = P$ satisfies it. A similar statement holds when $S = \mathcal{C}_{B_i}(W)$ (using Lemma 2.4), leaving us to deal with π_{m,B_i} .

Without loss of generality, we prove the result for the equation $(XB)^m X = P$. Assume that B and P are given and that X is a solution to $(XB)^m X = P$. Set $Y = B^{1/2} X B^{1/2}$, so that $X = B^{-1/2} Y B^{-1/2}$. Then

$$P = (B^{-1/2} Y B^{1/2})^m B^{-1/2} Y B^{-1/2} = B^{-1/2} Y^{m+1} B^{-1/2}.$$

Therefore, $Y^{m+1} = B^{1/2} P B^{1/2}$, from which it follows that Y is uniquely determined as $(B^{1/2} P B^{1/2})^{1/(m+1)}$. Hence, X must be the positive semidefinite matrix $B^{-1/2} (B^{1/2} P B^{1/2})^{1/(m+1)} B^{-1/2}$. Finally, substituting this X into the original equation does verify that it is a solution. This completes the proof. □

The shortest symmetric word equation without a known (closed-form) solution, as above, is $XBX^3BX = P$ (although it is uniquely solvable [21]). An exploration of which equations give rise to such explicit solutions is the focus of future work.

6. Reductions

The purpose of this section is to make some reductions that simplify the problem. Given the nature of Theorem 1.2 and Conjecture 1.3, we begin by noticing that we may assume that our interlaced symmetric words are of the following form:

$$S = X^{p_1} B_1 X^{p_2} B_2 \dots B_2 X^{p_2} B_1 X^{p_1}, \tag{4}$$

in which the exponents p_j are positive. This simplification is accomplished by observing, first, that powers of positive definite matrices are positive definite and, secondly, that congruences of positive semidefinite P are positive semidefinite.

We next establish that it suffices to verify our claims when P is invertible. We begin with a useful lemma.

LEMMA 6.1. *Let $p_1, \dots, p_k > 0$ and let B_1, \dots, B_{k-1} be positive definite matrices. Then, for any positive semidefinite matrix X , we have*

$$\ker X = \ker X^{p_k} B_{k-1} \dots B_2 X^{p_2} B_1 X^{p_1}.$$

Proof. Set $X = UDU^*$ for a unitary matrix U and $D = \text{diag}(\lambda_1, \dots, \lambda_n)$, in which $\lambda_1 \geq \dots \geq \lambda_n \geq 0$. Let $Y = X^{p_k} B_{k-1} \dots B_2 X^{p_2} B_1 X^{p_1}$, and notice that $\ker U^* X U = \ker U^* Y U$ if and only if $\ker X = \ker Y$. Thus, it suffices to argue that

$$\ker D = \ker D^{p_k} B_{k-1} \dots B_2 D^{p_2} B_1 D^{p_1},$$

whenever the B_i are positive definite matrices.

Let m be the largest integer such that $\lambda_m \neq 0$, and for each i , let \tilde{B}_i denote the $m \times m$ leading principal submatrix of B_i , which will be positive definite (see, for instance, [17, p. 472]). Additionally, set $\tilde{D} = \text{diag}(\lambda_1, \dots, \lambda_m)$. A straightforward block matrix multiplication then gives us that

$$D^{p_k} B_{k-1} \dots B_2 D^{p_2} B_1 D^{p_1} = \begin{bmatrix} \tilde{D}^{p_k} \tilde{B}_{k-1} \dots \tilde{B}_2 \tilde{D}^{p_2} \tilde{B}_1 \tilde{D}^{p_1} & 0 \\ 0 & 0 \end{bmatrix}. \tag{5}$$

Since the leading principal $m \times m$ matrix in this direct sum is invertible, the claim follows. \square

Using this lemma, we can prove the following reduction.

THEOREM 6.2. *If a symmetric word equation has a solution for every positive definite B_i and P , then the symmetric word equation has a solution for every positive definite B_i and positive semidefinite P .*

Proof. Performing a uniform unitary similarity, we may prove the theorem with the supposition that P is of the form,

$$\begin{bmatrix} \tilde{P} & 0 \\ 0 & 0 \end{bmatrix},$$

for a positive diagonal matrix \tilde{P} of rank m . Lemma 6.1 implies that any positive semidefinite solution X to the symmetric word equation $S(X, B_i) = P$ has the same block form as P . As in the lemma, let \tilde{B}_i denote the $m \times m$ leading principal (positive definite) submatrix of each B_i .

From these observations, it follows that positive semidefinite solutions X to the equation $S(X, B_i) = P$ correspond in a one-to-one manner with positive definite solutions \tilde{X} to the equation $S(\tilde{X}, \tilde{B}_i) = \tilde{P}$. This completes the proof. \square

The proof above also shows that the question of uniqueness found in Conjecture 1.3 may be simplified.

THEOREM 6.3. *If a symmetric word equation has a unique solution for all positive definite matrices B_i and P , then the symmetric word equation has a unique solution for all positive definite B_i and each positive semidefinite P .*

We close this section with an interesting interpretation of unique solvability.

PROPOSITION 6.4. *Fix positive definite matrices B_i in the unit ball, and an interlaced symmetric word $S(X, B_i)$ with equations that are uniquely solvable. Then the mapping $X \mapsto S(X, B_i)$ from the set of positive semidefinite matrices in the (closed) unit ball to its image is a homeomorphism.*

Proof. The assumptions imply that our map is bijective. Since the set of positive semidefinite matrices in the unit ball is compact, it follows that its inverse is also continuous. \square

7. Brouwer mapping degree

In this section, we give a brief overview of degree theory and some of its main implications. The bulk of this discussion is material taken from [7, 24, 32]. First we introduce some notation. Let U be a bounded open subset of \mathbb{R}^m . We denote the set of r -times differentiable functions from U to \mathbb{R}^m and from \bar{U} to \mathbb{R}^m by $C^r(U, \mathbb{R}^m)$ and $C^r(\bar{U}, \mathbb{R}^m)$, respectively (when $r = 0$, $C^r(U, \mathbb{R}^m)$ is the set of continuous functions). The *identity function* $\mathbb{1}$ satisfies $\mathbb{1}(\mathbf{x}) = \mathbf{x}$. If $f \in C^1(U, \mathbb{R}^m)$, then the *Jacobi matrix* of f at a point $\mathbf{x} \in U$ is

$$J_f(\mathbf{x}) = \left[\frac{\partial f_j}{\partial x_i}(\mathbf{x}) \right]_{1 \leq i, j \leq m}$$

and the *Jacobi determinant* (or simply *Jacobian*) of f at \mathbf{x} is

$$\det J_f(\mathbf{x}).$$

The set of *regular values* of f is

$$\text{RV}(f) = \{ \mathbf{y} \in \mathbb{R}^m : \forall \mathbf{x} \in f^{-1}(\mathbf{y}), J_f(\mathbf{x}) \neq 0 \}$$

and for $\mathbf{y} \in \mathbb{R}^m$, we set

$$D_{\mathbf{y}}^r(\bar{U}, \mathbb{R}^m) = \{ f \in C^r(\bar{U}, \mathbb{R}^m) : \mathbf{y} \notin f(\partial U) \}.$$

A function $\text{deg} : D_{\mathbf{y}}^0(\bar{U}, \mathbb{R}^m) \rightarrow \mathbb{R}$ which assigns to each $\mathbf{y} \in \mathbb{R}^m$ and $f \in D_{\mathbf{y}}^0(\bar{U}, \mathbb{R}^m)$ a real number $\text{deg}(f, U, \mathbf{y})$ will be called a *degree* if it satisfies the following conditions.

- (1) $\text{deg}(f, U, \mathbf{y}) = \text{deg}(f - \mathbf{y}, U, 0)$ (*translation invariance*).
- (2) $\text{deg}(\mathbb{1}, U, \mathbf{y}) = 1$ if $\mathbf{y} \in U$ (*normalization*).
- (3) If U_1 and U_2 are open, disjoint subsets of U such that $\mathbf{y} \notin f(\bar{U} \setminus (U_1 \cup U_2))$, then $\text{deg}(f, U, \mathbf{y}) = \text{deg}(f, U_1, \mathbf{y}) + \text{deg}(f, U_2, \mathbf{y})$ (*additivity*).
- (4) If $H(t) = tf + (1 - t)g \in D_{\mathbf{y}}^0(\bar{U}, \mathbb{R}^m)$ for all $t \in [0, 1]$, then $\text{deg}(f, U, \mathbf{y}) = \text{deg}(g, U, \mathbf{y})$ (*homotopy invariance*).

One should think of a degree map as somehow ‘counting’ the number of solutions to $f(\mathbf{x}) = \mathbf{y}$. Condition (1) reflects the fact that the solutions to $f(\mathbf{x}) = \mathbf{y}$ are the same as those of $f(\mathbf{x}) - \mathbf{y} = 0$, and since any multiple of a degree will satisfy (1) and (3), condition (2) is a normalization. Additionally, (3) is natural since it requires deg to be additive with respect to components. The following lemma gives a method to show the existence of solutions to $f(\mathbf{x}) = \mathbf{y}$ by calculating a degree.

LEMMA 7.1. *Suppose that $f \in D_{\mathbf{y}}^0(\bar{U}, \mathbb{R}^m)$. If a degree satisfies $\text{deg}(f, U, \mathbf{y}) \neq 0$, then $\mathbf{y} \in f(U)$.*

Proof. Using property (3) above with $U_1 = U$ and $U_2 = \emptyset$, we must have $\text{deg}(f, \emptyset, \mathbf{y}) = 0$. Again using (3) with $U_1 = U_2 = \emptyset$, it follows that if $\mathbf{y} \notin f(\bar{U})$ then $\text{deg}(f, U, \mathbf{y}) = 0$. The contrapositive is now what we want. \square

Of course, we need a theorem guaranteeing that a degree even exists.

THEOREM 7.2. *There is a unique degree deg . Moreover, $\text{deg}(\cdot, U, \mathbf{y}) : D_{\mathbf{y}}^0(\bar{U}, \mathbb{R}^m) \rightarrow \mathbb{Z}$.*

When functions are differentiable, the degree can be calculated explicitly in terms of Jacobians at solutions to the equation $f(\mathbf{x}) = \mathbf{y}$.

THEOREM 7.3. *Suppose that $f \in D_{\mathbf{y}}^1(\overline{U}, \mathbb{R}^m)$ and $\mathbf{y} \in \text{RV}$. Then the degree of f at \mathbf{y} with respect to U is given by*

$$\text{deg}(f, U, \mathbf{y}) = \sum_{\mathbf{x} \in f^{-1}(\mathbf{y})} \text{sgn det } J_f(\mathbf{x}),$$

where this sum is finite and we adopt the convention that $\sum_{\mathbf{x} \in \emptyset} = 0$.

The final property of the Brouwer degree that we will need is a stronger form of homotopy invariance than that provided by Property (4). We say that a function $H : \overline{U} \times [0, 1] \rightarrow \mathbb{R}^m$ is a C^0 homotopy between $f, g \in C^r(\overline{U}, \mathbb{R}^m)$ if H is continuous on $\overline{U} \times [0, 1]$ and if $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in \overline{U}$.

THEOREM 7.4. *Suppose that H is a C^0 homotopy between $f, g \in D_{\mathbf{y}}^0(\overline{U}, \mathbb{R}^m)$. Set $h_t(x) = H(x, t)$ and suppose that for each $t \in [0, 1]$, $h_t \in D_{\mathbf{y}}^0(\overline{U}, \mathbb{R}^m)$. Then $\text{deg}(f, U, \mathbf{y}) = \text{deg}(g, U, \mathbf{y})$.*

8. Estimates of solutions

This section is devoted to estimating the norms of solutions to symmetric word equations. In particular, we show that the set of positive semidefinite solutions to a fixed symmetric word equation $S(X, B_i) = P$ is bounded. Our estimate is the first step in a proof of Theorem 1.5. In what follows, we will be using the spectral norm [17, p. 295] on the set of $n \times n$ matrices, so that for positive semidefinite A , the norm of A is just the largest eigenvalue of A .

LEMMA 8.1. *Fix an interlaced symmetric word $S(X, B_i)$ and a number $\alpha \geq 1$. Then there exists a constant $C = C_{S, \alpha}$ depending only on S and α such that for all positive definite matrices B_i with $\|B_i\| \leq 1$ and $\|B_i^{-1}\| \leq \alpha$ and all positive semidefinite matrices P with $\|P\| \leq 1$, we have the estimate*

$$\|X\| \leq C \tag{6}$$

for any solution X of the word equation $S(X, B_i) = P$.

Proof. We proceed by way of contradiction. If the statement is false, then for each positive integer j there exist positive semidefinite matrices X_j, P_j and positive definite matrices $B_{i,j}$ such that $S(X_j, B_{i,j}) = P_j$, where $\|B_{i,j}\| \leq 1$, $\|B_{i,j}^{-1}\| \leq \alpha$, $\|P_j\| \leq 1$, and $\|X_j\| \geq j$. By taking a subsequence, if necessary, we may assume that there are positive semidefinite matrices B_i, P , and X such that $B_{i,j} \rightarrow B_i$, $P_j \rightarrow P$, and $\|X_j\|^{-1} X_j \rightarrow X$ as $j \rightarrow \infty$. It is clear that

$$\|X\| = 1. \tag{7}$$

Since $\|B_{i,j}^{-1}\|$ is bounded uniformly in j , each B_i is positive definite. Let s be the degree of S . Since $\|X_j\| \geq j$ for all j , if we let $j \rightarrow \infty$ in the equation

$$S(\|X_j\|^{-1} X_j, B_{i,j}) = \|X_j\|^{-s} P_j,$$

then it follows that

$$S(X, B_i) = 0.$$

Finally, an application of Lemma 6.1 gives $X = 0$, which contradicts (7) and finishes the proof. \square

Lemma 8.1 allows us to estimate $\|X\|$ in terms of the norms of the B_i and the norm of the word $S(X, B_i)$.

PROPOSITION 8.2. *Fix an interlaced symmetric word $S(X, B_1, \dots, B_k)$ of the form (4) with degree s and a number $\alpha \geq 1$. There exists a constant $C = C_{S, \alpha}$ depending only on S*

and α such that for all positive definite matrices B_i with $\|B_i\|\|B_i^{-1}\| \leq \alpha$ and any positive semidefinite X we have

$$\|X\| \leq C \|B_1\|^{-(2/s)} \dots \|B_k\|^{-(2/s)} \|S(X, B_i)\|^{1/s}. \tag{8}$$

Proof. Let $C = C_{S,\alpha}$ be the constant in Lemma 8.1. By Lemma 6.1, if $S(X, B_i) = 0$, then $X = 0$, and the bound is trivial. Otherwise, set $\tilde{B}_i = \|B_i\|^{-1} B_i$, $P = S(X, B_i)$, $\tilde{P} = \|P\|^{-1} P$, and $\tilde{X} = \|B_1\|^{2/s} \dots \|B_k\|^{2/s} \|P\|^{-(1/s)} X$. Since $\|\tilde{B}_i^{-1}\| = \|B_i\|\|B_i^{-1}\| \leq \alpha$ and $S(\tilde{X}, \tilde{B}_i) = \tilde{P}$, we may apply Lemma 8.1 to get

$$\|\tilde{X}\| \leq C. \tag{9}$$

Substituting $\tilde{X} = \|B_1\|^{2/s} \dots \|B_k\|^{2/s} \|P\|^{-(1/s)} X$ into (9) and rearranging produces (8). \square

9. Calculation of Jacobi matrices

From here on, we will assume that all matrices are real. We shall identify $\mathbb{M}_n = \mathbb{M}_n(\mathbb{R})$ with \mathbb{R}^d , where $d = n^2$, by means of the vec operator. If $A = [a_{ij}] \in \mathbb{M}_n$ then $\text{vec } A$ is the column vector obtained by stacking the columns of A below one another:

$$\text{vec } A = [a_{11} \dots a_{n1} \ a_{12} \dots \dots a_{nn}]^T.$$

Recall that the *Kronecker product* of two $n \times n$ matrices A and B is the matrix

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{n1}B & \cdots & a_{nn}B \end{bmatrix} \in \mathbb{M}_d.$$

The following lemma can be found in [25, p. 30]. We reproduce the proof below for the reader’s convenience.

LEMMA 9.1. *If $A, B, X \in \mathbb{M}_n$, then*

$$\text{vec}(AXB) = (B^T \otimes A)\text{vec } X.$$

Proof. For a given matrix Q , let Q_k denote the k th column of Q . Let $B = [b_{ij}]$. Then

$$\begin{aligned} (AXB)_k &= AXB_k \\ &= A \left(\sum_{i=1}^n b_{ik} X_i \right) = [b_{1k}A \dots b_{nk}A] \text{vec } X. \end{aligned}$$

Therefore,

$$\text{vec}(AXB) = \begin{bmatrix} b_{11}A & \cdots & b_{n1}A \\ \vdots & \ddots & \vdots \\ b_{1n}A & \cdots & b_{nn}A \end{bmatrix} \text{vec } X = (B^T \otimes A)\text{vec } X. \tag{10}$$

Suppose that $Y(X) \in \mathbb{M}_n$ is a function of the matrix variable $X \in \mathbb{M}_n$. Following [25], we define the *derivative* dY/dX of Y with respect to X to be the Jacobi matrix of $\text{vec } Y$ with respect to $\text{vec } X$. That is, if $[y_1, \dots, y_d]^T = \text{vec } Y$ and $[x_1, \dots, x_d]^T = \text{vec } X$, then

$$\frac{dY}{dX} = \left[\frac{\partial y_i}{\partial x_j} \right].$$

Notice that it follows from Lemma 9.1 that

$$\frac{d(AXB)}{dX} = B^T \otimes A. \tag{10}$$

Using (10), we derive a matrix calculus version of the product rule (see [25, 31] for more on matrix calculus).

PROPOSITION 9.2. *Let $Y(X) \in \mathbb{M}_n$ and $Z(X) \in \mathbb{M}_n$ be functions of the matrix variable $X \in \mathbb{M}_n$. Then*

$$\frac{d(YZ)}{dX} = (Z^T \otimes I) \frac{dY}{dX} + (I \otimes Y) \frac{dZ}{dX}. \tag{11}$$

Motivated by Theorem 7.3, we want to calculate the derivative dW/dX of a word $W = W(X, B_1, \dots, B_k)$. To state the result, we need to introduce some notation. Let W have degree $s \geq 1$. Enumerate the occurrences of X in $W(X, B_i)$ from left to right, and for each $j \in \{1, \dots, s\}$ let $W_j^L(X, B_i)$ be the portion of $W(X, B_i)$ that appears to the left of the j th occurrence of X . For instance, if

$$W(X, B_1, B_2) = B_2^3 X B_1^2 B_2 X B_2 B_1 X^2 B_2 X,$$

then $W_4^L(X, B_1, B_2) = B_2^3 X B_1^2 B_2 X B_2 B_1 X$. We adopt the convention that $W_1^L = I$ if X is the first letter of the word. In a similar way we define $W_j^R(X, B_i)$ to be the portion of $W(X, B_i)$ that appears to the right of the j th appearance of X . Notice that

$$W(X, B_i) = W_j^L(X, B_i) X W_j^R(X, B_i)$$

for any $j = 1, \dots, s$.

PROPOSITION 9.3. *Let $W = W(X, B_i)$ be a word of degree s and $B_i \in \mathbb{M}_n$. Then*

$$\frac{dW}{dX} = \sum_{j=1}^s W_j^R(X, B_i)^T \otimes W_j^L(X, B_i). \tag{12}$$

Proof. We proceed by induction on the length of W . For the words X and BX ($B = B_1, \dots, B_k$), (12) is a special case of (10). Now suppose that (12) holds for a fixed word $W = W(X, B_i)$ of degree s . Pick $B \in \{B_1, \dots, B_k\}$ and set

$$\overline{W}(X, B_i) = W(X, B_i)B.$$

Then (10) and (11) imply that

$$\begin{aligned} \frac{d\overline{W}}{dX} &= (B^T \otimes I) \frac{dW}{dX} \\ &= (B^T \otimes I) \sum_{j=1}^s W_j^R(X, B_i)^T \otimes W_j^L(X, B_i) \\ &= \sum_{j=1}^s (W_j^R(X, B_i)B)^T \otimes W_j^L(X, B_i) \\ &= \sum_{j=1}^s \overline{W}_j^R(X, B_i)^T \otimes \overline{W}_j^L(X, B_i), \end{aligned}$$

so formula (12) holds for \overline{W} .

Next set $\widetilde{W}(X, B_i) = W(X, B_i)X$. Appealing again to (10) and (11), we compute:

$$\begin{aligned} \frac{d\widetilde{W}}{dX} &= (X^T \otimes I) \frac{dW}{dX} + (I \otimes W(X, B_i)) \\ &= (X^T \otimes I) \sum_{j=1}^s W_j^R(X, B_i)^T \otimes W_j^L(X, B_i) + (I \otimes \widetilde{W}_{s+1}^L) \\ &= \sum_{j=1}^{s+1} \widetilde{W}_j^R(X, B_i)^T \otimes \widetilde{W}_j^L(X, B_i), \end{aligned}$$

and so (12) holds for \widetilde{W} . This completes the induction and the proof. □

We next write down expression (12) for some explicit interlaced symmetric words, beginning with the most basic one.

EXAMPLE 9.4. For a positive integer s , the Jacobi matrix of $\text{vec } X^s$ with respect to $\text{vec } X$ is given by

$$\frac{dX^s}{dX} = \sum_{j=1}^s (X^{s-j})^T \otimes X^{j-1}. \tag{13}$$

In particular, since the Kronecker product of two positive (semi)definite matrices is also positive (semi)definite (see [18, p. 245]), dX^s/dX is positive (semi)definite whenever X is positive (semi)definite.

EXAMPLE 9.5. Consider the symmetric word S in two letters given by

$$S(X, B) = XBX^2B^3X^2BX.$$

If B is positive definite and X is symmetric, then

$$\begin{aligned} \frac{dS}{dX} &= XBX^2B^3X^2B \otimes I + XBX^2B^3X \otimes XB + XBX^2B^3 \otimes XBX \\ &\quad + XBX \otimes XBX^2B^3 + XB \otimes XBX^2B^3X + I \otimes XBX^2B^3X^2B. \end{aligned}$$

10. The Brouwer degree of symmetric word equations

Our proof of Theorem 1.5 consists of two main steps. In the first, we calculate the degree of the simple map $\varphi_s(X) = X^s$ and show that it is 1. In the second, we create a homotopy from the function $f(X) = S(X, B_i)$ to $\varphi_s(X)$ and apply Theorem 7.4. Before initiating our proof, we need to identify the set of real positive definite matrices with an open set in Euclidean space. To this end, we identify the set Sym_n of real symmetric matrices with \mathbb{R}^m , in which $m = \frac{1}{2}n(n + 1)$, by identifying a real symmetric matrix $A = [a_{ij}]$ with the point

$$\mu(A) = (a_{11}, \dots, a_{n1}, a_{22}, \dots, a_{n2}, \dots, a_{nn}).$$

More precisely, if $A \in \mathbb{M}_n$ then we define $\mu(A) = (y_1, \dots, y_m)$, where

$$y_{\frac{1}{2}(2n-j)(j-1)+i} = a_{ij}, \quad 1 \leq j \leq i \leq n.$$

The restriction $\mu|_{\text{Sym}_n}$ is a linear isomorphism from Sym_n onto \mathbb{R}^m . We denote by ν the inverse of $\mu|_{\text{Sym}_n}$. Let

$$\mathcal{O} = \{\mu(X) : X \text{ is positive definite}\}.$$

The set of positive definite matrices is therefore identified with the open subset $\mathcal{O} \subset \mathbb{R}^m$, and the set of positive semidefinite matrices is identified with the set $\overline{\mathcal{O}}$.

Define a function $\tilde{\varphi}_s : \mathbb{R}^m \rightarrow \mathbb{R}^m$ by

$$\tilde{\varphi}_s = \mu \circ \varphi_s \circ \nu.$$

Since φ_s maps Sym_n into itself, it follows that $\tilde{\varphi}_s(\mu(X)) = \mu(X^s)$ for every symmetric matrix X . We intend to show that $\det J_{\tilde{\varphi}_s}(\mu(X)) > 0$ when X is positive definite. First, however, we need a lemma describing a relationship between eigenvalues of Jacobi matrices for functions $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and their restrictions \tilde{f} to certain subspaces. In what follows, the set of eigenvalues of a matrix H is denoted by $\sigma(H)$.

LEMMA 10.1. *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ be a C^1 map, and let $V \subseteq \mathbb{R}^d$ be a linear subspace of \mathbb{R}^d such that $f(V) \subseteq V$. Let $\pi : \mathbb{R}^m \rightarrow V$ be a linear isomorphism, and let $\tilde{f} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ be given by $\tilde{f} = \pi^{-1} \circ f \circ \pi$. Then for every $\mathbf{x} \in V$, we have*

$$\sigma(J_{\tilde{f}}(\pi^{-1}(\mathbf{x}))) \subseteq \sigma(J_f(\mathbf{x})).$$

In particular, if $\mathbf{x} \in V$ and $J_f(\mathbf{x})$ is positive definite, then $J_{\tilde{f}}(\pi^{-1}(\mathbf{x}))$ has positive eigenvalues.

Proof. Let $\{e_1, \dots, e_d\}$ be the standard basis for \mathbb{R}^d . By choosing a linear change of variables $u : \mathbb{R}^d \rightarrow \mathbb{R}^d$ such that $u(V) = \text{span}\{e_1, \dots, e_m\}$ and considering the C^1 map $g = u \circ f \circ u^{-1}$, we may reduce to the case that $V = \text{span}\{e_1, \dots, e_m\}$. We may likewise assume that $\pi(z_1, \dots, z_m) = (z_1, \dots, z_m, 0, \dots, 0)$.

Write $f = (f_1, \dots, f_d)$ and let $\mathbf{x} \in V$. If $j \leq m < k$, then we have

$$f_k(\mathbf{x} + te_j) = 0 \quad \forall t \in \mathbb{R}$$

since $f(V) \subseteq V$. Therefore,

$$\frac{\partial f_k}{\partial x_j}(\mathbf{x}) = 0 \quad \forall j \leq m < k. \tag{14}$$

It follows that $J_f(\mathbf{x})$ has the block form

$$J_f(\mathbf{x}) = \begin{bmatrix} J_0 & * \\ 0 & * \end{bmatrix},$$

in which J_0 is the $m \times m$ leading principle submatrix of $J_f(\mathbf{x})$. In particular, this implies that $\sigma(J_0) \subset \sigma(J_f(\mathbf{x}))$. It is straightforward to verify that

$$\tilde{f}(x_1, \dots, x_m) = (f_1(x_1, \dots, x_m, 0, \dots, 0), \dots, f_m(x_1, \dots, x_m, 0, \dots, 0)),$$

from which it follows that $J_{\tilde{f}}(\pi^{-1}(\mathbf{x})) = J_0$. This proves the lemma. □

LEMMA 10.2. *At any positive definite matrix X , the Jacobi matrix $J_{\tilde{\varphi}_s}(\mu(X))$ of the map $\tilde{\varphi}_s$ has positive eigenvalues. In particular, $\det J_{\tilde{\varphi}_s}(\mu(X)) > 0$.*

Proof. Let $V = \{\text{vec } X : X \in \text{Sym}_n\}$ be the linear subspace of \mathbb{R}^d identified with Sym_n . The function $\phi_s : \mathbb{R}^d \rightarrow \mathbb{R}^d$ defined by

$$\phi_s(\text{vec } X) = \text{vec } X^s$$

maps V into itself. Let $\pi = \text{vec} \circ \nu$, and notice that $\pi : \mathbb{R}^m \rightarrow V$ is a linear isomorphism and that $\tilde{\varphi}_s = \pi^{-1} \circ \phi_s \circ \pi$. By Example 9.4, if X is a positive definite matrix, then $J_{\phi_s}(\text{vec } X) = dX^s/dX$ is also positive definite. Applying Lemma 10.1, we conclude that the Jacobi matrix $J_{\tilde{\varphi}_s}(\pi^{-1}(\text{vec } X)) = J_{\tilde{\varphi}_s}(\mu(X))$ has positive eigenvalues at any positive definite X . □

Our next result verifies Theorem 1.5 for the special case $S(X, B_i) = X^s$.

PROPOSITION 10.3. *Let s be a positive integer, P a positive definite matrix, and $\mathcal{V} \subset \mathcal{O}$ a bounded open set containing $\mu(P^{1/s})$. Let g be the function $\tilde{\varphi}_s$ restricted to \mathcal{O} . Then*

$$\deg(g, \mathcal{V}, \mu(P)) = 1.$$

Proof. Lemma 10.2 implies that $\mu(P)$ is a regular value for g . Using Theorem 7.3 and Lemma 10.2, we calculate that

$$\deg(g, \mathcal{V}, \mu(P)) = \sum_{\mathbf{x} \in g^{-1}(\mu(P))} \operatorname{sgn} \det J_g(\mathbf{x}) = \operatorname{sgn} \det J_g(\mu(P^{1/s})) = 1. \quad \square$$

The following straightforward fact will be used in our proof of Theorem 1.5.

LEMMA 10.4. *Let V be the set of positive definite matrices of norm less than K , which is greater than 0. Then the boundary of V is given by*

$$\partial V = \{X \in \bar{V} : \det(X) = 0\} \cup \{X \in \bar{V} : \|X\| = K\}.$$

We are now ready to calculate the Brouwer degree of a general symmetric word equation; the boundedness part of Theorem 1.5 follows from Lemma 8.1.

Proof of Theorem 1.5. From the discussion in Section 6, we may assume that our interlaced symmetric word $S(X, B_1, \dots, B_k)$ is of the form (4). Fix positive definite matrices B_1, \dots, B_k and a positive definite matrix P , and set $f(X) = S(X, B_i)$. Also set $\tilde{f} = \mu \circ f \circ \nu$. We will show that there is a bounded, open subset $\mathcal{V} \subset \mathcal{O}$ such that

$$\deg(\tilde{f}, \mathcal{V}, \mu(P)) = 1. \tag{15}$$

Let $t \in [0, 1]$. By Proposition 8.2, there exists a constant K independent of t such that any positive semidefinite solution X of the equation $S(X, tB_i + (1 - t)I) = P$ has $\|X\| < K$. Indeed, if

$$\alpha = \max_{1 \leq i \leq k, 0 \leq t \leq 1} \|tB_i + (1 - t)I\| \cdot \|(tB_i + (1 - t)I)^{-1}\| < \infty$$

and

$$\beta = \min_{1 \leq i \leq k, 0 \leq t \leq 1} \|tB_i + (1 - t)I\| > 0,$$

then we must have

$$\|X\| \leq C_{S,\alpha} \beta^{-(2k/s)} \|P\|^{1/s} < \infty.$$

Let $V = V_K$ be the open set of positive definite matrices with norm less than K . For each $t \in [0, 1]$, let f_t be the function from the positive semidefinite matrices into itself given by $f_t(X) = S(X, tB_i + (1 - t)I)$. From our choice of K , it follows that $f_t(X) \neq P$ when X is positive definite with $\|X\| = K$. Moreover, if X is singular, then taking a determinant shows that $f_t(X) \neq P$. Thus $P \notin f_t(\partial V)$ for each $t \in [0, 1]$ by Lemma 10.4.

Set $\mathcal{V} = \mu(V)$ and $\tilde{f}_t = \mu \circ f_t \circ \nu$. Then we have $\mu(P) \notin \tilde{f}_t(\partial \mathcal{V})$ for $t \in [0, 1]$. Since $(\mathbf{x}, t) \mapsto \tilde{f}_t(\mathbf{x})$ is continuous, Theorem 7.4 implies that

$$\deg(\tilde{f}_0, \mathcal{V}, \mu(P)) = \deg(\tilde{f}_1, \mathcal{V}, \mu(P)).$$

Since $\tilde{f}_0 = \tilde{\varphi}_s$ and $\tilde{f}_1 = \tilde{f}$, equation (15) now follows from Proposition 10.3. □

11. Nonuniqueness of symmetric word equations

The following corollary of Theorem 1.5 is a crucial ingredient in our proof of Theorem 1.4.

COROLLARY 11.1. Fix an interlaced symmetric word S and let $\tilde{f} = \tilde{f}_S$ be as in the proof of Theorem 1.5. Suppose that there is a positive definite matrix X_0 such that

$$\det J_{\tilde{f}}(\mu(X_0)) < 0.$$

Then the symmetric word equation

$$S(X, B_i) = S(X_0, B_i)$$

has at least two real solutions X .

Proof. Let X_0 be as in the statement of the corollary, and set $P = S(X_0, B_i)$. If $\mu(P)$ is a regular value of \tilde{f} , then Theorems 1.5 and 7.3 imply that there must be at least two solutions X_1 and X_2 of $S(X, B_i) = P$ such that

$$\det J_{\tilde{f}}(\mu(X_i)) > 0, \quad i = 1, 2.$$

If $\mu(P)$ is not a regular value of \tilde{f} , then there exists a positive definite matrix X_1 such that $S(X_1, B_i) = P$ and

$$J_{\tilde{f}}(\mu(X_1)) = 0.$$

Since $J_{\tilde{f}}(\mu(X_1)) \neq J_{\tilde{f}}(\mu(X_0))$, it follows that $X_0 \neq X_1$. □

Let S and \tilde{f} be as in Corollary 11.1. We outline a method for obtaining the smaller Jacobian matrix $J_{\tilde{f}}(\mu(X))$ from the larger Jacobian matrix dS/dX . To simplify the bookkeeping of indices, define

$$\alpha(i, j) = n(j - 1) + i, \quad i, j = 1, \dots, n$$

and

$$\beta(k, l) = \frac{1}{2}(2n - l)(l - 1) + k, \quad 1 \leq l \leq k \leq n.$$

Thus if $X = [x_{ij}] \in \mathbb{M}_n$, then the $\alpha(i, j)$ th entry of $\text{vec } X$ is equal to x_{ij} , $i, j = 1, \dots, n$. Likewise, the $\beta(k, l)$ th entry of $\mu(X)$ is x_{kl} , $1 \leq l \leq k \leq n$.

The Jacobi matrix $J_{\tilde{f}}$ of the map

$$\begin{aligned} \tilde{f} &= \mu \circ (X \mapsto S(X, B_i)) \circ \nu \\ &= (\mu \circ \text{vec}^{-1}) \circ (\text{vec} \circ (X \mapsto S(X, B_i)) \circ \text{vec}^{-1}) \circ (\text{vec} \circ \nu) \end{aligned}$$

is given by

$$J_{\tilde{f}}(\mu(X)) = M \left(\frac{dS}{dX} \right) N,$$

in which $M \in \mathbb{M}_{m \times d}$ is the matrix representation of $\mu \circ \text{vec}^{-1}$ and $N \in \mathbb{M}_{d \times m}$ is the matrix representation of $\text{vec} \circ \nu$. It is easy to see that if $1 \leq i, j \leq n$ and $1 \leq l \leq k \leq n$, then the $(\alpha(i, j), \beta(k, l))$ entry of N is

$$\begin{cases} 1 & \text{if } i = k, j = l \text{ or } i = l, j = k, \\ 0 & \text{otherwise,} \end{cases}$$

and the $(\beta(k, l), \alpha(i, j))$ entry of M is

$$\begin{cases} 1 & \text{if } i = k, j = l, \\ 0 & \text{otherwise.} \end{cases}$$

EXAMPLE 11.2. When $n = 2$, the matrices M and N as described above are:

$$M = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad N = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

We are now ready to prove the main result of this section.

Proof of Theorem 1.4. Let A_1 and B_1 be as in Section 4, and let S be the symmetric word

$$S(X, B) = XBX^2B^3X^2BX.$$

Let $f(X) = S(X, A_1)$ and $\tilde{f} = \mu \circ f \circ \nu$. Using MAPLE, we calculate that

$$\det J_{\tilde{f}}(B_1) = -633705909477329213831177437148144640 < 0.$$

(Code that performs this calculation and the one found in Theorem 11.4 is available at <http://math.berkeley.edu/~sarm> or <http://www.math.tamu.edu/~chillar>.)

By Corollary 11.1, it follows that the symmetric word equation

$$S(X, A_1) = S(B_1, A_1)$$

has at least two distinct real positive definite solutions X . □

REMARK 11.3. Despite many numerical attempts using generalized Newton methods, we were unable to produce any of these other solutions. It appears that the ill-conditioning of A_1 and B_1 necessary to achieve a negative determinant causes difficulty for numerical equation solvers.

We note that there are many other words which can be shown to have multiple solutions using the techniques found in the proof of Theorem 1.4. We list a few of them below:

$$\begin{aligned} XBX^kBX, & \quad 9 \leq k \leq 20, \\ XBXB^2X^k B^2XBX, & \quad 2 \leq k \leq 16, \\ XBX^k B^3X^kBX, & \quad 2 \leq k \leq 15, \\ XB^2XBX^k BXB^2X, & \quad 6 \leq k \leq 40. \end{aligned}$$

In general, we do not know how to characterize those equations which give rise to unique solutions.

The techniques that we have developed here can also be used to show unique solvability of some particular symmetric word equations. As an illustration, we close this section by proving that the word equation $S(X, B) = XBX^2B^3X^2BX = P$ is uniquely solvable in 2×2 real positive definite matrices. From Example 9.5, we have

$$\begin{aligned} \frac{dS}{dX} &= XBX^2B^3X^2B \otimes I + XBX^2B^3X \otimes XB + XBX^2B^3 \otimes XBX \\ &\quad + XBX \otimes XBX^2B^3 + XB \otimes XBX^2B^3X + I \otimes XBX^2B^3X^2B, \end{aligned} \tag{16}$$

in which I is the 2×2 identity matrix and X and B are 2×2 positive definite matrices.

Theorem 1.5 and the remarks in this section reduce unique solvability to a verification of whether $\det[M(dS/dX)N] > 0$ for all 2×2 real positive definite matrices X and B . This observation is used to prove the following theorem.

THEOREM 11.4. *The word equation $S(X, B) = XBX^2B^3X^2BX = P$ has a unique real solution for all 2×2 real positive definite matrices B and P .*

Proof. It is clear that in the theorem statement we may assume that B is diagonal (perform a uniform unitary similarity) and that X has determinant 1 (homogeneity of (16)). We may, therefore, parameterize X and B as

$$X = \begin{bmatrix} x & y \\ y & \frac{1+y^2}{x} \end{bmatrix}, \quad B = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}; \quad a, b, x > 0.$$

With the assistance of MAPLE, the expression $\det[M(dS/dX)N]$ factors into a product of terms that are easily seen to be positive:

$$\begin{aligned} & \frac{12a^5b^5}{x^5} (3a^5x^{10} + 3b^5y^{10} + 15b^5y^8 + 30b^5y^4 + 30b^5y^6 + 15b^5y^2 + 3a^4bx^8 \\ & + 22a^2b^3x^4y^4 + 12a^2b^3x^4y^6 + 15a^2b^3x^2y^2 + 27a^2b^3x^2y^4 + 21a^2b^3x^2y^6 + 6a^2b^3x^2y^8 \\ & + 5ab^4x^6y^2 + 3b^5 + 6a^5x^8y^2 + 3a^5x^6y^2 + 3a^5x^6y^4 + 3a^2b^3x^2 + 3ab^4x^2 + 6b^5x^4y^4 \\ & + 18b^5x^2y^6 + 3b^5y^2x^4 + 6b^5y^2x^2 + 18b^5y^4x^2 + 3b^5y^6x^4 + 6b^5y^8x^2 + 3a^3b^2x^8 \\ & + 8ab^4x^4y^2 + 6a^3x^8y^2b^2 + 10a^3x^6y^2b^2 + 12a^3x^6y^4b^2 + 7a^3x^4y^2b^2 + 13a^3x^4y^4b^2 \\ & + 6a^3x^4y^6b^2 + 3ax^6y^4b^4 + 14ax^4y^4b^4 + 6ax^4y^6b^4 + 12ax^2y^2b^4 + 18ax^2y^4b^4 \\ & + 12ax^2y^6b^4 + 3ax^2y^8b^4 + 3a^4bx^8y^2 + 8a^4bx^6y^2 + 6a^4bx^6y^4 + 5a^4bx^4y^2 + 8a^4bx^4y^4 \\ & + 3a^4bx^4y^6 + 7a^2b^3x^6y^2 + 6a^2b^3x^6y^4 + 10a^2b^3x^4y^2). \end{aligned}$$

This completes the proof. □

We close with a conjecture that vastly generalizes this last result.

CONJECTURE 11.5. Symmetric word equations in 2×2 positive definite matrices have unique solutions.

REMARK 11.6. Standard transformations for 2×2 matrices reduce the general problem to the real case. Therefore, for instance, Theorem 11.4 is fully general in the sense of this conjecture.

Acknowledgements. The authors would like to thank Bill Helton for several interesting conversations about this problem, and Robert McCann for his helpful advice and references.

References

1. T. ANDO, 'Concavity of certain maps on positive definite matrices and applications to Hadamard products', *Linear Algebra Appl.* 26 (1979) 203–241.
2. T. ANDO, C.-K. LI and R. MATHIAS, 'Geometric means', *Linear Algebra Appl.* 385 (2004) 305–334.
3. D. BESSIS, P. MOUSSA and M. VILLANI, 'Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics', *J. Math. Phys.* 16 (1975) 2318–2325.
4. D. C. DOWSON and B. V. LANDAU, 'The Frechet distance between multivariate normal distributions', *J. Multivariate Anal.* 12 (1982) 450–455.
5. M. DRMOTA, W. SCHACHERMAYER and J. TEICHMANN, 'A hyper-geometric approach to the BMV-conjecture', *Monatsh. Math.* 146 (2005) 179–201.
6. M. FANNES and D. PETZ, 'Perturbation of Wigner matrices and a conjecture', *Proc. Amer. Math. Soc.* 131 (2003) 1981–1988.
7. I. FONSECA and W. GANGBO, 'Degree theory in analysis and applications' (Oxford University Press, New York, 1995).
8. C. R. GIVENS and R. M. SHORTT, 'A class of Wasserstein metrics for probability distributions', *Michigan Math. J.* 31 (1984) 231–240.
9. F. HANSEN, 'Trace functions as Laplace transforms', *J. Math. Phys.* 47 (2006) 043504.
10. R. HAUSER and Y. LIM, 'Self-scaled barriers for irreducible symmetric cones', *SIAM J. Optim.* 12 (2002) 715–723.

11. C. HILLAR, 'Advances on the Bessis–Moussa–Villani trace conjecture', *Linear Algebra Appl.* 426 (2007) 130–142.
12. C. HILLAR and C. R. JOHNSON, 'Eigenvalues of words in two positive definite letters', *SIAM J. Matrix Anal. Appl.* 23 (2002) 916–928.
13. C. HILLAR and C. R. JOHNSON, 'Symmetric word equations in two positive definite letters', *Proc. Amer. Math. Soc.* 132 (2004) 945–953.
14. C. HILLAR and C. R. JOHNSON, 'On the positivity of the coefficients of a certain polynomial defined by two positive definite matrices', *J. Stat. Phys.* 118 (2005) 781–789.
15. C. HILLAR, C. R. JOHNSON and I. M. SPITKOVSKY, 'Positive eigenvalues and two-letter generalized words', *Electron. J. Linear Algebra* 9 (2002) 21–26.
16. C. HILLAR and C. R. JOHNSON, 'Positive eigenvalues of generalized words in two Hermitian positive definite matrices', *Novel approaches to hard discrete optimization* (eds. P. Pardalos and H. Wolkowicz), Fields Institute Communications 37 (2003) 111–122.
17. R. HORN and C. R. JOHNSON, *Matrix analysis* (Cambridge University Press, New York, 1985).
18. R. HORN and C. R. JOHNSON, *Topics in matrix analysis* (Cambridge University Press, New York, 1991).
19. M. KNOTT and C. S. SMITH, 'On the optimal mappings of distributions', *J. Optim. Theory Appl.* 43 (1984) 39–49.
20. J. LAWSON and Y. LIM, 'The geometric mean, matrices, and more', *Amer. Math. Monthly* 108 (2001) 797–812.
21. J. LAWSON and Y. LIM, 'Solving symmetric matrix word equations via symmetric space machinery', *Linear Algebra Appl.* 414 (2006) 560–569.
22. E. H. LIEB and R. SEIRINGER, 'Equivalent forms of the Bessis–Moussa–Villani conjecture', *J. Stat. Phys.* 115 (2004) 185–190.
23. Y. LIM, 'Applications of geometric means on symmetric cones', *Math. Ann.* 319 (2001) 457–468.
24. N. LLOYD, *Degree theory* (Cambridge University Press, London, 1978).
25. J. R. MAGNUS and M. NEUDECKER, *Matrix differential calculus with applications in statistics and econometrics* (John Wiley, New York, 1999).
26. R. J. MCCANN, 'A convexity principle for interacting gases', *Adv. Math.* 128 (1997) 153–179.
27. R. J. MCCANN and A. M. OBERMAN, 'Exact semi-geostrophic flows in an elliptical ocean basin', *Nonlinearity* 17 (2004) 1891–1922.
28. N. MILLER, '3 × 3 cases of the Bessis–Moussa–Villani conjecture', Senior Thesis, Princeton University, Princeton, NJ, 2004.
29. Y. E. NESTEROV and M. J. TODD, 'Self-scaled barriers and interior-point methods for convex programming', *Math. Oper. Res.* 22 (1997) 1–42.
30. I. OLKIN and F. PUKELSHEIM, 'The distance between two random vectors with given dispersion matrices', *Linear Algebra Appl.* 48 (1982) 257–263.
31. P. SEBASTIANI, 'On the derivatives of matrix powers', *SIAM J. Matrix Anal. Appl.* 17 (1996) 640–648.
32. G. TESCHL, *Nonlinear functional analysis*, 2005, <http://www.mat.univie.ac.at/~gerald/ftp/book-nlfa/index.html>.

Scott N. Armstrong
 Department of Mathematics
 University of California
 Berkeley
 CA 94720
 USA

sarm@math.berkeley.edu

Christopher J. Hillar
 Department of Mathematics
 Texas A&M University
 College Station
 TX 77843
 USA

chillar@math.tamu.edu