J Stat Phys (2010) 139: 779–799
DOI 10.1007/s10955-010-9959-y

Sum-of-Squares Results for Polynomials Related to the Bessis–Moussa–Villani Conjecture

Benoît Collins · Kenneth J. Dykema · Francisco Torres-Ayala

Received: 18 May 2009 / Accepted: 1 April 2010 / Published online: 21 April 2010

Abstract We show that the polynomial $S_{m,k}(A, B)$, that is the sum of all words in noncommuting variables A and B having length m and exactly k letters equal to B, is not equal to a sum of commutators and Hermitian squares in the algebra $\mathbf{R}\langle X, Y\rangle$, where $X^2 = A$ and $Y^2 = B$, for all even values of m and k with $6 \leq k \leq m - 10$, and also for $(m, k) = (12, 6)$. This leaves only the case $(m, k) = (16, 8)$ open. This topic is of interest in connection with the Lieb–Seiringer formulation of the Bessis–Moussa–Villani conjecture, which asks whether $\mathrm{Tr}(S_{m,k}(A, B)) \geq 0$ holds for all positive semidefinite matrices A and B. These results eliminate the possibility of using "descent + sum-of-squares" to prove the BMV conjecture.

We also show that $S_{m,4}(A, B)$ is equal to a sum of commutators and Hermitian squares in $\mathbf{R}\langle A, B\rangle$ when m is even and not a multiple of 4, which implies $\mathrm{Tr}(S_{m,4}(A, B)) \geq 0$ holds for all Hermitian matrices A and B, for these values of m.

Keywords BMV conjecture · Hermitian squares

B. Collins research supported in part by NSERC grant RGPIN/341303-2007 and by the Granma and Galoisint ANR grants.
K.J. Dykema research supported in part by NSF grants DMS-0600814 and DMS-0901220.

B. Collins
Department of Mathematics and Statistics, University of Ottawa, 585 King Edward, Ottawa, ON K1N 6N5, Canada
e-mail: bcollins@uottawa.ca

B. Collins
CNRS, Department of Mathematics, Lyon 1 Claude Bernard University, Vileurbanne, France

K.J. Dykema (✉) · F. Torres-Ayala
Department of Mathematics, Texas A&M University, College Station, TX 77843-3368, USA
e-mail: kdykema@math.tamu.edu

F. Torres-Ayala
e-mail: francisc@math.tamu.edu

1 Introduction

While working on quantum statistical mechanics, Bessis, Moussa and Villani [1] conjectured in 1975 that for any positive semidefinite Hermitian matrices A and B, the function $t \mapsto \mathrm{Tr}(e^{A-tB})$ is the Laplace transform of a positive measure supported in $\mathbf{R}_+$. This is referred to as the Bessis–Moussa–Villani or BMV conjecture. In 2004, Lieb and Seiringer [9] proved that the BMV conjecture is equivalent to the following reformulation: for every A and B as above, all of the coefficients of the polynomial

$$p(t) = \mathrm{Tr}((A+tB)^m) \in \mathbf{R}[t] \tag{1}$$

are nonnegative. Recently, there has been much activity around this algebraic reformulation, (see [2, 4–6, 8]). The latest state of knowledge is summarized in [6], and we'll review this here.

Let $S_{m,k}(A,B)$ denote the sum of all words of length m in A and B having k letters equal to B and $m-k$ equal to A. Thus, the coefficient of t^k in the polynomial $p(t)$ of (1) is equal to the trace of $S_{m,k}(A,B)$, and the Lieb–Seiringer reformulation of the BMV conjecture is that this trace is always nonnegative. An important result, due to Hillar [4], is that if this conjecture fails for some (m,k), then it fails for all (m',k') satisfying $k' \geq k$ and $m'-k' \geq m-k$. We'll refer to this as Hillar's descent theorem.

One strategy that has been used to show that the trace of $S_{m,k}(A,B)$ is nonnegative for certain values of m and k is to let X and Y be formal square roots of A and B, respectively and, working in the algebra $\mathbf{R}\langle X,Y\rangle$ of polynomials in noncommuting variables X and Y, to show that $S_{m,k}(A,B)$ is equal to a sum of commutators $[g,h]=gh-hg$ and Hermitian squares f^*f. Here, the algebra $\mathbf{R}\langle X,Y\rangle$ is endowed with the involutive $*$-operation that is anti-multiplicative and so that $X=X^*$ and $Y=Y^*$ are Hermitian. We adopt the notation of [6] and say that two elements $a,b \in \mathbf{R}\langle X,Y\rangle$ are *cyclically equivalent* (written $a \overset{\mathrm{cyc}}{\sim} b$) if they differ by a sum of commutators. We will use repeatedly Proposition 2.3 of [6], which states that two words v and w in X and Y are cyclically equivalent if and only if they can be written $v=u_1u_2$ and $w=u_2u_1$ for words u_1 and u_2 in X and Y, and that two polynomials $a,b \in \mathbf{R}\langle X,Y\rangle$ are cyclically equivalent if and only if for each cyclic equivalence class $[w]$ of words in X and Y, the sum over all v in $[w]$ of the coefficients a_v of a agrees with the sum over all v in $[w]$ of the coefficients b_v of b. It is clear that any element of $\mathbf{R}\langle A,B\rangle$ that is cyclically equivalent in $\mathbf{R}\langle X,Y\rangle$ to a sum of Hermitian squares in $\mathbf{R}\langle X,Y\rangle$ must have nonnegative trace whenever A and B are replaced by positive semidefinite matrices, and this has been the strategy used to show that $S_{m,k}(A,B)$ has nonnegative trace, for certain values of m and k. We will adopt the terminology of [6] and write Θ^2 to denote the set of elements of $\mathbf{R}\langle X,Y\rangle$ that are cyclically equivalent to sums of Hermitian squares in $\mathbf{R}\langle X,Y\rangle$. (It is not difficult to see that $\Theta^2_{\mathbf{C}} \cap \mathbf{R}\langle X,Y\rangle = \Theta^2$, where $\Theta^2_{\mathbf{C}}$ is the analogous quantity in $\mathbf{C}\langle X,Y\rangle$.)

Clearly, $S_{m,k}(A,B) \in \Theta^2$ if and only if $S_{m,m-k}(A,B) \in \Theta^2$. Due to work of Hägele [3], Landweber and Speer [8], Burgdorf [2] and Klep and Schweighofer [6], it is known that $S_{m,k}(A,B) \in \Theta^2$ holds

- whenever $k \in \{0,1,2,4\}$
- for $m=14$ and $k=6$
- for $m \in \{7,11\}$ and $k=3$

These cases together with Hillar's descent theorem implied that the Lieb–Seiringer formulation of the BMV-conjecture holds for $m \leq 13$ (see [6]). On the other hand, it is known that $S_{m,k}(A,B) \notin \Theta^2$ holds

- whenever $m \geq 12$ or $m \in \{6, 8, 9, 10\}$ and $k = 3$
- whenever $m \geq 10$ and $5 \leq k \leq m - 5$ and either k or m is odd.

It was hoped that proofs of $S_{m,k}(A, B) \in \Theta^2$ for other values of m and k would be possible, so as to prove the conjecture for more values of m, and possibly even to prove the BMV conjecture itself.

These results left open the cases $(m, k) = (12, 6)$ and $m \geq 16$, $6 \leq k \leq m - 6$ with both m and k even. In this paper (see Sect. 2), we prove $S_{m,k}(A, B) \notin \Theta^2$ whenever m and k are even and $6 \leq k \leq m - 10$. Using $S_{m,k}(A, B) = S_{m,m-k}(B, A)$, this leaves open only the cases $(m, k) = (12, 6)$ and $(m, k) = (16, 8)$. We resolve the first of these cases by showing, via an easier argument, $S_{12,6}(A, B) \notin \Theta^2$. The case of $(m, k) = (16, 8)$ remains open, though, as indicated in [6], numerical evidence seems to suggest it does not lie in Θ^2.

Our results, thus, show that it is impossible to prove the BMV conjecture by showing that $S_{m,k}(A, B)$ is cyclically equivalent to a sum of Hermitian squares for sufficiently many values of m and k. However there are other plausible approaches to showing $\mathrm{Tr}(S_{m,k}(A, B)) \geq 0$ must always hold.

Though our proofs are straightforward and easy to check by hand, to find them we calculated with Mathematica 7.0 [10], on an Apple MacBook running OS X version 10.4.11.

While exploring, we found (see Proposition 3.3) that if m is even and is not a multiple of 4, then $S_{m,4}(A, B)$ is equal to a sum of commutators and Hermitian squares in $\mathbf{R}\langle A, B\rangle$. Thus, we do not need the square roots of A and B: for these values of m we have $\mathrm{Tr}(S_{m,4}(A, B)) \geq 0$ whenever A and B are Hermitian matrices.

Question 1.1 Do we have $\mathrm{Tr}(S_{m,k}(A, B)) \geq 0$ whenever A and B are Hermitian matrices and m and k are even integers, $m \geq k$?

Using Hillar's descent theorem, a positive answer to Question 1.1 would imply the Lieb–Seiringer formulation of the BMV conjecture.

We will prove the following theorem in Sect. 4. It shows that Question 1.1 has an equivalent formulation that seems easier to satisfy, and is analogous to Theorem 1.10 of [4]. Note that $S_{m,k}(A, B)$ is Hermitian whenever A and B are Hermitian.

Theorem 1.2 *Fix $n, m, k \in \mathbf{N}$ with m and k even and $m \geq k$. Then the following are equivalent*:

1. *for all $n \times n$ Hermitian matrices A and B, we have* $\mathrm{Tr}(S_{m,k}(A, B)) \geq 0$,
2. *for all $n \times n$ Hermitian matrices A and B, either $S_{m,k}(A, B) = 0$ or $S_{m,k}(A, B)$ has a strictly positive eigenvalue.*

In Sect. 3 we also show (Proposition 3.8) that $S_{8,4}(A, B)$ is not cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle A, B\rangle$. This makes the case $(m, k) = (8, 4)$ of particular interest for Question 1.1.

Our interest in Question 1.1 has two motivations. One is its relation to the BMV conjecture. Although the question is known to be stronger than the BMV conjecture and we have no particular reason to think it will be easier to prove than the BMV conjecture itself, it is clearly related to the BMV conjecture and it may be helpful to explore it. A second motivation is the relation to Connes' embedding problem. For positive semidefinite matrices A and B, the trace of $S_{6,3}(A, B)$ is always nonnegative, though it is not cyclically equivalent to a sum of squares in $\mathbf{C}\langle X, Y\rangle$; as was pointed out in [7], this makes $S_{6,3}(A, B)$, with A and B positive operators in a II_1-factor, an interesting test case for Connes' embedding problem.

In a similar way, if Question 1.1 turns out to have a positive answer for $S_{8,4}(A,B)$, then because of Proposition 3.8, then it will provide another interesting test case for Connes' embedding problem, involving self-adjoint operators. At this point, it seems important to generate such test cases.

After a first version of this paper was circulated, we learned that S. Burgdorf (see Remarks (b) and (c) of Sect. 4 of [2]) had, long previously to us, also found that if m is not a multiple of 4, then $S_{m,4}(A,B)$ is cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle A,B\rangle$; no proof was given in [2].

2 Some Non-sum-of-Squares Results

In this section, we show that $S_{m,k}(A,B)$ is not cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle X,Y\rangle$ for various values of m and k, all of which are even.

Let $W_{q,p}(A,B)$ denote the set of all words in A and B containing q A's and p B's. Let Z denote the column vector whose entries are all words in $W_{\ell,k}(A,B)$ in some fixed order, and similarly let Z_X and, respectively, Z_Y be column vectors containing all elements of $XW_{\ell-1,k}(A,B)X$, respectively, $YW_{\ell,k-1}(A,B)Y$. Klep and Schweighofer have shown (Proposition 3.3 of [6]) that, for integers k and ℓ, $S_{2(k+\ell),2k}(A,B)$ is cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle X,Y\rangle$ if and only if there are real, positive semidefinite matrices H, H_X and H_Y such that

$$Z^*HZ+Z_X^*H_XZ_X+Z_Y^*H_YZ_Y\overset{\text{cyc}}{\sim}S_{2(k+\ell),2k}(A,B), \tag{2}$$

where Z^* denotes the row vector whose entries are the adjoints of the entries of Z, *etc*. Let us denote the matrix entry of H corresponding to words $u,v\in W_{\ell,k}(A,B)$ by $H(u,v)$, and similarly for H_X and H_Y. Thus, we have

$$Z^*HZ=\sum_{u,v\in W_{\ell,k}(A,B)}H(u,v)u^*v, \tag{3}$$

and similarly for the other two terms.

Remark 2.1 If H is a matrix as appearing in (3), and if $\widehat{H}$ is the matrix defined by $\widehat{H}(u,v)=H(u^*,v^*)$, then

$$Z^*\widehat{H}Z=\sum_{u,v}H(u^*,v^*)u^*v=\sum_{u,v}H(u,v)uv^*\overset{\text{cyc}}{\sim}\sum_{u,v}H(v,u)v^*u=Z^*HZ,$$

where the last equality uses that H is symmetric. In a similar way, defining $\widehat{H}_X(u,v)=\widehat{H}_X(u^*,v^*)$ and $\widehat{H}_Y(u,v)=\widehat{H}_Y(u^*,v^*)$, we have

$$Z_X^*\widehat{H}_XZ_X\overset{\text{cyc}}{\sim}Z_X^*H_XZ_X,$$

$$Z_Y^*\widehat{H}_YZ_Y\overset{\text{cyc}}{\sim}Z_Y^*H_YZ_Y.$$

Consequently, if H, H_X and H_Y are such that (2) holds, then by replacing H with $(H+\widehat{H})/2$, if necessary, and similarly for H_X and H_Y, we may without loss of generality assume

$$H(u,v)=H(u^*,v^*),\quad (u,v\in W_{\ell,k}(A,B)), \tag{4}$$

$$H_X(u,v) = H_X(u^*,v^*), \quad (u,v \in XW_{\ell-1,k}(A,B)X), \tag{5}$$

$$H_Y(u,v) = H_Y(u^*,v^*), \quad (u,v \in YW_{\ell,k-1}(A,B)Y). \tag{6}$$

Suppose, furthermore, we have $k=\ell$. Let σ is the map on words that exchanges A and B and exchanges X and Y, extended by linearity to $\mathbf{R}\langle X,Y\rangle$. Then $\sigma(Z^*HZ) = Z^*H^\sigma Z$, where $H^\sigma(u,v) = H(\sigma(u),\sigma(v))$, and, similarly, $\sigma(Z_X^*H_XZ_X) = Z_Y^*H_X^\sigma Z_Y$ and $\sigma(Z_Y^*H_YZ_Y) = Z_X^*H_Y^\sigma Z_X$, where $H_X^\sigma(u,v) = H_X(\sigma(u),\sigma(v))$ and $H_Y^\sigma(u,v) = H_Y(\sigma(u),\sigma(v))$. Consequently, if H, H_X and H_Y are such that (2) holds, then since $S_{2(k+\ell),2k}(A,B)$ is σ-invariant and since σ respects $\overset{\text{cyc}}{\sim}$, by replacing H with $(H+H^\sigma)/2$, H_X with $(H_X+H_Y^\sigma)/2$ and H_Y with $(H_Y+H_X^\sigma)/2$, if necessary, we may without loss of generality assume

$$H(\sigma(u),\sigma(v)) = H(u,v), \quad (u,v \in W_{\ell,k}(A,B)), \tag{7}$$

$$H_Y(\sigma(u),\sigma(v)) = H_X(u,v), \quad (u,v \in XW_{\ell-1,k}(A,B)X). \tag{8}$$

Since $\sigma(u^*) = \sigma(u)^*$, we can assume that (4)–(6) and (7)–(8) hold simultaneously.

We note that the relation (4) will be used in this section, while (7) will be used only in the proof of Proposition 3.8, and the conditions on H_X and H_Y won't be needed at all in this paper.

Remark 2.2 For a given word $w \in W_{2\ell,2k}(A,B)$, we are interested in the different ways we can have

$$w \overset{\text{cyc}}{\sim} u^*v, \quad (u,v \in W_{\ell,k}(A,B)), \tag{9}$$

$$w \overset{\text{cyc}}{\sim} u_X^*v_X, \quad (u_X,v_X \in XW_{\ell-1,k}(A,B)X), \tag{10}$$

$$w \overset{\text{cyc}}{\sim} u_Y^*v_Y, \quad (u_Y,v_Y \in YW_{\ell,k-1}(A,B)Y). \tag{11}$$

Indeed, if $|[w]|$ denotes the number of different elements of $W_{2\ell,2k}(A,B)$ that are cyclically equivalent to w, and assuming (2) holds, then we have

$$\begin{aligned}|[w]| = &\sum_{\{(u,v)|u^*v \overset{\text{cyc}}{\sim} w\}} H(u,v) + \sum_{\{(u_X,v_X)|u_X^*v_X \overset{\text{cyc}}{\sim} w\}} H_X(u_X,v_X)\\ &+ \sum_{\{(u_Y,v_Y)|u_Y^*v_Y \overset{\text{cyc}}{\sim} w\}} H_Y(u_Y,v_Y),\end{aligned} \tag{12}$$

where the respective sums are over all pairs (u,v) such that (9) holds, all pairs (u_X,v_X) such that (10) holds and all pairs (u_Y,v_Y) such that (11) holds. To find all the ways we have (9), we can write down all the cyclic permutations of w and record those for which the first $k+\ell$ letters consists of ℓ A's and k B's. Furthermore, if we have an instance of (10) with $u_X = Xu'X$ and $v_X = Xv'X$, $u',v' \in W_{\ell-1,k}(A,B)$, then $w \overset{\text{cyc}}{\sim} X(u')^*Av'X \overset{\text{cyc}}{\sim} A(u')^*Av'$; this yields an instance of (9), where both u^* and v start with A, and clearly each such instance corresponds in this manner to an instance of (10). Similarly, the instances of (11) are in one-to-one correspondence with those instances of (9) where both u^* and v start with B.

We will apply (in a finite dimensional setting) the following elementary lemma, whose proof we provide for completeness.

Lemma 2.3 *Let $\mathcal{H} = \mathcal{H}_1 \oplus \mathcal{H}_2$ be an orthogonal direct sum decomposition of a Hilbert space and let $T \in B(\mathcal{H})$ be a positive operator: $T \geq 0$. With respect to the given decomposition of $\mathcal{H}$, write T in block form*

$$T = \begin{pmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{pmatrix},$$

where $T_{ij} : \mathcal{H}_j \to \mathcal{H}_i$. Suppose $v \in \ker T_{11} \subseteq \mathcal{H}_1$. Then $v \in \ker T_{21}$.

Proof If $T_{21}v \neq 0$, then there is $w \in \mathcal{H}_2$ such that $\langle T_{21}v, w\rangle < 0$. Letting $t > 0$ and using $T_{12} = T_{21}^*$, we have

$$\langle T(v \oplus tw), v \oplus tw\rangle = 2t \operatorname{Re}\langle T_{21}v, w\rangle + t^2\langle T_{22}w, w\rangle. \tag{13}$$

But taking t small enough forces the right-hand-side of (13) to be negative, which contradicts $T \geq 0$. □

Proposition 2.4 *Let k and ℓ be integers, $k \geq 3$ and $\ell \geq 5$. Then $S_{2(\ell+k),2k}(A, B)$ is not cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle X, Y\rangle$.*

Proof Suppose the contrary, to obtain a contradiction. Let H, H_X and H_Y be real, positive semidefinite matrices so that (2) holds, and without loss of generality assume also the property (4) in Remark 2.1 holds.

We consider five elements of $W_{2\ell,2k}(A, B)$ and the different ways of writing them as in (9). These elements are

$$\begin{aligned} w_1 &= A^{2\ell}B^{2k}, & w_2 &= A^{2\ell-2}B^{k-1}A^2B^{k+1}, \\ w_3 &= A^{\ell+1}B^2A^{\ell-1}B^{2k-2}, & w_4 &= A^{2\ell-4}B^{k-1}A^2B^2A^2B^{k-1}, \\ w_5 &= A^{\ell-1}B^2A^{\ell-1}B^{k-1}A^2B^{k-1}, \end{aligned}$$

and their factorizations will be in terms of the elements

$$\begin{aligned} u_1 &= A^{\ell}B^{k}, & v_1 &= u_1^* = B^kA^{\ell}, \\ u_2 &= A^{\ell-2}B^{k-1}A^2B, & v_2 &= u_2^* = BA^2B^{k-1}A^{\ell-2}, \\ u_3 &= AB^2A^{\ell-1}B^{k-2}, & v_3 &= u_3^* = B^{k-2}A^{\ell-1}B^2A, \\ u_4 &= AB^{k-1}A^{\ell-1}B, & v_4 &= u_4^* = BA^{\ell-1}B^{k-1}A \end{aligned}$$

of $W_{\ell,k}(A, B)$. Note that these are all distinct if $k \geq 4$; in the case $k = 3$, the six elements $u_1, u_2, u_3, v_1, v_2, v_3$ are distinct but we have $u_4 = u_3$ and $v_4 = v_3$. This will not bother us.

We begin with the easiest of the w_j to factorize, namely, w_1. In the Table 1 are listed all the cyclically equivalent forms of w_1 and it is indicated which of these can be factored as in (9).

This also shows that there are no factorizations as in (10) or (11) (see Remark 2.2). Since w_1 has $2(k+\ell)$ cyclically equivalent forms, by (12) we must have $H(u_1, u_1) + H(v_1, v_1) = 2(k+\ell)$. Since we have $H(v_1, v_1) = H(u_1, u_1)$, we get

$$H(u_1, u_1) = k + \ell. \tag{14}$$

Table 1 Forms of $w_1 = A^{2\ell}B^{2k}$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^{2k}A^{2\ell-j}$	$(1 \le j \le 2\ell)$	$j=\ell$	$v_1^*v_1$
$B^jA^{2\ell}B^{2k-j}$	$(1 \le j \le 2k)$	$j=k$	$u_1^*u_1$

Table 2 Forms of $w_2 = A^{2\ell-2}B^{k-1}A^2B^{k+1}$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^{k-1}A^2B^{k+1}A^{2\ell-2-j}$	$(1 \le j \le 2\ell-2)$	$j=\ell-2$	$v_2^*v_1$
$B^jA^2B^{k+1}A^{2\ell-2}B^{k-1-j}$	$(1 \le j \le k-1)$		none
$A^jB^{k+1}A^{2\ell-2}B^{k-1}A^{2-j}$	$(1 \le j \le 2)$		none
$B^jA^{2\ell-2}B^{k-1}A^2B^{k+1-j}$	$(1 \le j \le k+1)$	$j=k$	$u_1^*u_2$

Table 3 Forms of $w_3 = A^{\ell+1}B^2A^{\ell-1}B^{2k-2}$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^2A^{\ell-1}B^{2k-2}A^{\ell+1-j}$	$(1 \le j \le \ell+1)$	$j=1$	$v_3^*v_1$
$B^jA^{\ell-1}B^{2k-2}A^{\ell+1}B^{2-j}$	$(1 \le j \le 2)$		none
$A^jB^{2k-2}A^{\ell+1}B^2A^{\ell-1-j}$	$(1 \le j \le \ell-1)$		none
$B^jA^{\ell+1}B^2A^{\ell-1}B^{2k-2-j}$	$(1 \le j \le 2k-2)$	$j=k$	$u_1^*u_3$

Table 4 Forms of $w_4 = A^{2\ell-4}B^{k-1}A^2B^2A^2B^{k-1}$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^{k-1}A^2B^2A^2B^{k-1}A^{2\ell-4-j}$	$(1 \le j \le 2\ell-4)$	$j=\ell-2$	$v_2^*v_2$
$B^jA^2B^2A^2B^{k-1}A^{2\ell-4}B^{k-1-j}$	$(1 \le j \le k-1)$		none
$A^jB^2A^2B^{k-1}A^{2\ell-4}B^{k-1}A^{2-j}$	$(1 \le j \le 2)$		none
$B^jA^2B^{k-1}A^{2\ell-4}B^{k-1}A^2B^{2-j}$	$(1 \le j \le 2)$	$j=1$	$u_2^*u_2$
$A^jB^{k-1}A^{2\ell-4}B^{k-1}A^2B^2A^{2-j}$	$(1 \le j \le 2)$		none
$B^jA^{2\ell-4}B^{k-1}A^2B^2A^2B^{k-1-j}$	$(1 \le j \le k-1)$		none

The cyclically equivalent forms and all factorizations of w_2, w_3, w_4 and w_5 as in (9) are given in Tables 2–5. (Note that the assertions in rows 2, 3 and 6 of Table 4 do require $\ell \ge 5$.)

From these, we see that each of the words w_j, $2 \le j \le 5$ has $2(k+\ell)$ different cyclically equivalent forms, and none have factorizations involving X or Y, as in (10) or (11). Looking at the two factorizations of w_2, and using (12) and $H(v_2, v_1) = H(v_1, v_2) = H(u_1, u_2)$, we conclude

$$H(u_1, u_2) = k + \ell. \tag{15}$$

Similarly, considering all the factorizations of w_3, w_4 and w_5 we get, respectively,

$$H(u_1, u_3) = k + \ell \tag{16}$$

$$H(u_2, u_2) = k + \ell \tag{17}$$

$$2H(u_2, u_3) + H(u_4, u_4) = k + \ell. \tag{18}$$

Table 5 Forms of $w_5 = A^{\ell-1}B^2A^{\ell-1}B^{k-1}A^2B^{k-1}$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^2A^{\ell-1}B^{k-1}A^2B^{k-1}A^{\ell-1-j}$	$(1\le j\le \ell-1)$	$j=1$	$v_3^*v_2$
$B^jA^{\ell-1}B^{k-1}A^2B^{k-1}A^{\ell-1}B^{2-j}$	$(1\le j\le 2)$	$j=1$	$u_4^*u_4$
$A^jB^{k-1}A^2B^{k-1}A^{\ell-1}B^2A^{\ell-1-j}$	$(1\le j\le \ell-1)$	$j=\ell-2$	$v_2^*v_3$
$B^jA^2B^{k-1}A^{\ell-1}B^2A^{\ell-1}B^{k-1-j}$	$(1\le j\le k-1)$	$j=1$	$u_2^*u_3$
$A^jB^{k-1}A^{\ell-1}B^2A^{\ell-1}B^{k-1}A^{2-j}$	$(1\le j\le 2)$	$j=1$	$v_4^*v_4$
$B^jA^{\ell-1}B^2A^{\ell-1}B^{k-1}A^2B^{k-1-j}$	$(1\le j\le k-1)$	$j=k-2$	$u_3^*u_2$

Table 6 Forms of $w_6 = A^6B^6$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^6A^{6-j}$	$(1\le j\le 6)$	$j=3$	$v_5^*v_5$
$B^jA^6B^{6-j}$	$(1\le j\le 6)$	$j=3$	$u_5^*u_5$

Now from (14)–(17), for the 3×3 submatrix of H corresponding to the entries u_1, u_2, u_3, we have

$$\begin{pmatrix} H(u_1,u_1) & H(u_1,u_2) & H(u_1,u_3)\\ H(u_1,u_2) & H(u_2,u_2) & H(u_2,u_3)\\ H(u_1,u_3) & H(u_2,u_3) & H(u_3,u_3)\end{pmatrix} = \begin{pmatrix} k+\ell & k+\ell & k+\ell\\ k+\ell & k+\ell & H(u_2,u_3)\\ k+\ell & H(u_2,u_3) & H(u_3,u_3)\end{pmatrix}. \tag{19}$$

From (19), the positivity of H and Lemma 2.3, we obtain also $H(u_2,u_3)=k+\ell$. But then, from (18), we must have $H(u_4,u_4) = -(k+\ell)$, which contradicts the positive semidefiniteness of H. □

Proposition 2.5 *$S_{12,6}(A,B)$ is not cyclically equivalent to a sum of squares in $\mathbf{R}\langle X,Y\rangle$.*

Proof This is like the proof of Proposition 2.4, but easier. Again we assume, to obtain a contradiction, that H, H_X and H_Y are real, positive semidefinite matrices such that (2) holds (with $k=\ell=3$) and that the properties (9)–(11) hold. We need only consider the words

$$w_6 = A^6B^6, \qquad w_7 = A^4B^2A^2B^4, \qquad w_8 = A^2B^2A^2B^2A^2B^2$$

in $W_{6,6}(A,B)$ and their factorizations, which will be in terms of the elements

$$u_5 = A^3B^3, \qquad v_5 = u_5^* = B^3A^3,$$
$$u_6 = AB^2A^2B, \qquad v_6 = u_6^* = BA^2B^2A$$

of $W_{3,3}(A,B)$. These factorizations are given in Tables 6–8.

Again, w_6, w_7 and w_8 have no factorizations as in (10) or (11). From Table 6, we see that w_6 has 12 distinct cyclically equivalent forms, and since $H(u_5,u_5)=H(v_5,v_5)$, from (12) we get $H(u_5,u_5)=6$. From Table 7 and $H(v_6,v_5)=H(u_6,u_5)=H(u_5,u_6)$, we get $H(u_5,u_6)=6$, while from Table 8 we see that w_8 has only four distinct cyclically equivalent forms, and we get $H(u_6,u_6)=2$. The 2×2 submatrix of H corresponding to $\{u_5,u_6\}$ is, therefore,

$$\begin{pmatrix} H(u_5,u_5) & H(u_5,u_6)\\ H(u_6,u_5) & H(u_6,u_6)\end{pmatrix} = \begin{pmatrix} 6 & 6\\ 6 & 2\end{pmatrix},$$

Table 7 Forms of $w_7 = A^4B^2A^2B^4$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^2A^2B^4A^{4-j}$	$(1 \le j \le 4)$	$j=1$	$v_6^*v_5$
$B^jA^2B^4A^4B^{2-j}$	$(1 \le j \le 2)$		none
$A^jB^4A^4B^2A^{2-j}$	$(1 \le j \le 2)$		none
$B^jA^4B^2A^2B^{4-j}$	$(1 \le j \le 4)$	$j=3$	$u_5^*u_6$

Table 8 Forms of $w_8 = A^2B^2A^2B^2A^2B^2$ and factorizations as in (9)

cyclically equivalent form		j value	factorization
$A^jB^2A^2B^2A^2B^2A^{2-j}$	$(1 \le j \le 2)$	$j=1$	$v_6^*v_6$
$B^jA^2B^2A^2B^2A^2B^{2-j}$	$(1 \le j \le 2)$	$j=1$	$u_6^*u_6$

which is not positive semidefinite. This gives a contradiction. □

3 Sums of Squares in $\mathbf{R}\langle A, B\rangle$

In this section, we prove some results related to Question 1.1. As per the discussion in the introduction (see Proposition 2.3 of [6]), we say $f, g \in \mathbf{R}\langle A, B\rangle$ are cyclically equivalent if and only if $f - g$ is a sum of commutators of elements from $\mathbf{R}\langle A, B\rangle$. This holds if and only if, for every word w in A and B, the sum over words v that are cyclic permutations of w of the coefficients in f of v agrees with the same sum for g.

Clearly, if $S_{m,k}(A, B)$ is cyclically equivalent to a sum $\sum_i f_i^* f_i$ of Hermitian squares, for $f_i \in \mathbf{R}\langle A, B\rangle$, then Question 1.1 has a positive answer for this particular pair (m, k).

Of course, $S_{2m,0}(A, B) = A^{2m}$ is a Hermitian square in $\mathbf{R}\langle A, B\rangle$, for every integer $m \ge 0$.

Verification of the following two lemmas is straightforward.

Lemma 3.1 *Let $m \in \mathbf{N}$. Then*

$$S_{4m,2}(A, B) \overset{\text{cyc}}{\sim} m f_m^* f_m + 2m \sum_{j=0}^{m-1} f_j^* f_j,$$

where

$$f_0 = BA^{2m-1},$$

$$f_j = A^{j-1}BA^{2m-j} + A^jBA^{2m-j-1}, \quad (1 \le j \le m).$$

Lemma 3.2 *Let $m \in \mathbf{N}$. Then*

$$S_{4m+2,2}(A, B) \overset{\text{cyc}}{\sim} (2m+1) \sum_{j=0}^{m} f_j^* f_j,$$

where

$$f_0 = BA^{2m},$$

$$f_j = A^{j-1}BA^{2m-j+1} + A^jBA^{2m-j}, \quad (1 \le j \le m).$$

The next proposition shows that $S_{2q,4}(A,B)$ is cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle A,B\rangle$, when q is odd. Note that Klep and Schweighofer in Sect. 5 of [6] proved this in the case $q=7$. In fact, we found the expression (20) below by exploration using Mathematica [10] and checked it by computation for all values of $m \le 20$. The best proof we could find, which is given below, turned out to be surprisingly intricate.

Proposition 3.3 *Let $m \in \mathbf{N}$. Then*

$$S_{4m+2,4}(A,B) \overset{\text{cyc}}{\sim} (2m+1)\sum_{p=0}^{m} f_p^* f_p, \tag{20}$$

where

$$\begin{aligned} f_0 &= \sum_{s=0}^{2m-1} BA^{2m-s-1}BA^s, \\ f_p &= \sum_{i=p-1}^{p}\sum_{s=p}^{2m-i-1} A^iBA^{2m-s-i-1}BA^s, \quad (1 \le p \le m-1), \\ f_m &= A^{m-1}B^2A^m. \end{aligned}$$

As before $W_{q,4}(A,B)$ denotes the set of all words in A and B with exactly q A's and four B's. Let $\mathbf{N}_0 = \mathbf{N} \cup \{0\}$. For $\iota = (\iota_1, \iota_2, \iota_3, \iota_4.\iota_5) \in \mathbf{N}_0^5$ let

$$E(\iota) = A^{\iota_1}BA^{\iota_2}BA^{\iota_3}BA^{\iota_4}BA^{\iota_5}$$

and take

$$I = \{\iota \in \mathbf{N}_0^5 \mid \iota_1+\iota_2+\iota_3+\iota_4+\iota_5 = 4m-2\}.$$

Note that the map $\iota \mapsto E(\iota)$ gives a bijection from I onto $W_{4m-2,4}(A,B)$. With this notation we may write

$$S_{4m+2,4}(A,B) = \sum_{\iota \in I} E(\iota).$$

The proof of Proposition 3.3 will use the following three lemmas. The first of these is readily verified, and a proof will be omitted.

Lemma 3.4 *Each word in $W_{4m-2,4}(A,B)$ is cyclically equivalent to a unique word of the form*

$$BA^{k_1}BA^{k_2}BA^{k_3}BA^{k_4},$$

where $\kappa = (0,k_1,k_2,k_3,k_4) \in I$ satisfies either

$$k_1 \le k_3 \quad \textit{and} \quad k_2 < k_4 \tag{21}$$

or

$$k_1 = k_3 \quad \le \quad k_2 = k_4. \tag{22}$$

We will call the words (or indices) described in (21) and (22) *canonically ordered* and those of the form (21) will be called *type I* while those given by (22) will be called *type II*. Since the first letter of any canonically ordered word is a B, canonically ordered words are parameterized by only four non-negative integers, and we'll frequently omit to write the first element of a canonically ordered index κ, since it is always zero.

Lemma 3.5

$$\#\{\kappa \in I \mid \kappa \text{ is canonically ordered of type I}\} = \frac{2m(2m-1)(2m+1)}{3}.$$

$$\#\{\kappa \in I \mid \kappa \text{ is canonically ordered of type II}\} = m.$$

Proof We recall that a partition of $n \in \mathbf{N}$ into k parts is a k-tuple $(a_1, a_2, \dots, a_k)$ such that $1 \le a_1 \le a_2 \le \cdots \le a_k$ and $a_1 + a_2 + \cdots + a_k = n$. We denote it as $(a_1, a_2, \dots, a_k) \vdash n$.

Consider the sets

$$B = \{(a, b, a_1, a_2, b_1, b_2) \in \mathbf{N}^6 \mid a + b = 4m + 1, (a_1, a_2) \vdash a, (b_1, b_2) \vdash b\}$$

and

$$A = \{\kappa \in I \mid \kappa \text{ is canonically ordered of type I}\}.$$

Take the function from A into B given by

$$(k_1, k_2, k_3, k_4) \mapsto (k_1 + k_3 + 2, k_2 + k_4 + 1, k_1 + 1, k_3 + 1, k_2 + 1, k_4).$$

One can show this function is a bijection onto B. Thus,

$$\#A = \sum_{\substack{(a,b)\in\mathbf{N}^2 \\ a+b=4m+1}} \left\lfloor \frac{a}{2} \right\rfloor \left\lfloor \frac{b}{2} \right\rfloor = \frac{2}{3} m(2m-1)(2m+1).$$

Similarly, the function

$$(k_1, k_2, k_3, k_4) \mapsto (k_1 + 1, k_2 + 1)$$

is a bijection from $\{\kappa \in I \mid \kappa \text{ is canonically ordered of type II}\}$ onto the set $\{(a, b) \in \mathbf{N}^2 \mid (a, b) \vdash 2m + 1\}$. Hence

$$\#\{\kappa \in I \mid \kappa \text{ is canonically ordered of type II}\} = \left\lfloor \frac{2m+1}{2} \right\rfloor = m.$$ □

The following lemma is easily verified by writing out the cyclically equivalent forms of words; see Tables 1–8 for other exercises of this sort.

Lemma 3.6 *Let $w \in W_{4m-2,4}(A, B)$ be a canonically ordered word. If w is of type I, then there are $4m + 2$ words in $W_{4m-2,4}(A, B)$ that are cyclically equivalent to w, while if w is of type II, then there are $2m + 1$ words in $W_{4m-2,4}(A, B)$ that are cyclically equivalent to w.*

Proof of Proposition 3.3 Let $l = 2m - 1$.

For $g \in \mathbf{R}\langle A, B\rangle$ and w a word in A and B, we let $c_w(g)$ denote the coefficient of w in g. By Lemmas 3.4 and 3.6 it will suffice to show, for every canonically ordered word $w \in W_{4m-2,4}(A, B)$,

$$\sum_{\{v \mid v \overset{\text{cyc}}{\sim} w\}} \sum_{p=0}^{m} c_v(f_p^* f_p) = \begin{cases} 2, & w \text{ of type I,} \\ 1, & w \text{ of type II} \end{cases} \tag{23}$$

i.e., for each such w, there is only one representative in $\sum_{p=0}^{m} f_p^* f_p$ if w is type II and exactly two representatives if w is type I.

We begin by taking a closer look at each $f_p^* f_p$. We have

$$f_0^* f_0 = \sum_{0 \le s,t \le l} A^s B A^{l-s} B^2 A^{l-t} B A^t = \sum_{\iota \in I_0} E(\iota),$$

where

$$I_0 = \{\iota = (s, l-s, 0, l-t, t) \mid 0 \le s, t \le l\}$$

and for $1 \le p \le m-1$,

$$\begin{aligned} f_p^* f_p &= \sum_{p-1 \le i,j \le p} \sum_{\substack{p \le s \le l-i \\ p \le t \le l-j}} A^s B A^{l-i-s} B A^{i+j} B A^{l-j-t} B A^t \\ &= \sum_{\iota \in I_p(p-1,p-1)} E(\iota) + \sum_{\iota \in I_p(p-1,p)} E(\iota) + \sum_{\iota \in I_p(p,p-1)} E(\iota) + \sum_{\iota \in I_p(p,p)} E(\iota), \end{aligned}$$

where

$$I_p(i, j) = \{\iota = (s, l-i-s, i+j, l-j-t, t) \mid p \le s \le l-i, p \le t \le l-j\},$$

while

$$f_m^* f_m = \sum_{\iota \in I_m} E(\iota),$$

where

$$I_m = \{(m, 0, 2m-2, 0, m)\}.$$

We also write $I_0(0, 0) = I_0$ and $I_m(m-1, m-1) = I_m$.

Let J be the disjoint union

$$J_0 \sqcup \left(\bigsqcup_{p=1}^{m-1} \bigsqcup_{p-1 \le i,j \le p} J_p(i, j) \right) \sqcup J_m,$$

where each $J_p(i, j)$ is a copy of the corresponding $I_p(i, j)$ and similarly for $J_0 = J_0(0, 0)$ and $J_m = J_m(m-1, m-1)$. Formally, given $0 \le p \le m$ and $\max\{0, p-1\} \le i, j \le \min\{p, m-1\}$, we set

$$J_p(i, j) = \{(p, i, j, \iota) \mid \iota \in I_p(i, j)\}$$

and we let $\alpha_p^{(i,j)} : I_p(i, j) \to J_p(i, j)$ be the bijection given by $\iota \mapsto (p, i, j, \iota)$.

Consider the function $O : I \to I$, where $O(\iota)$ is the index of the canonically ordered word that is cyclically equivalent to $E(\iota)$. This function O is explicitly given on I_0 and on each $I_p(i,j)$ $(1 \le p \le m-1,\ p-1 \le i, j \le p)$ as follows. For $\iota = (s, l-i-s, i+j, l-j-t, t) \in I_p(i,j)$ we have

$$O(\iota) = \begin{cases} U(i,j,s,t), & \text{if } (i=j \text{ and } t>s) \text{ or } (i>j \text{ and } t-1>s) \\ & \qquad \text{or } (j>i \text{ and } t>s-1), \\ L(i,j,s,t), & \text{if } (i=j \text{ and } t\le s) \text{ or } (i>j \text{ and } t\le s-1) \\ & \qquad \text{or } (j>i \text{ and } t\le s-1), \end{cases}$$

where U and L are given by

$$U(i,j,s,t) = (0,l,0,l) + \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 \\ -1 & 0 & -1 & 0 \end{pmatrix} \begin{pmatrix} i \\ j \\ s \\ t \end{pmatrix},$$

$$L(i,j,s,t) = (l,0,l,0) + \begin{pmatrix} -1 & 0 & -1 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{pmatrix} \begin{pmatrix} i \\ j \\ s \\ t \end{pmatrix}.$$

The canonical form of an element of J is naturally taken to be the same as the canonical form of the element of I to which it corresponds and we denote the "canonical form map" also by $O : J \to I$.

We now work on proving (23). For $0 \le p \le m-1$ define

$$\iota_p = (p, l-2p, 2p, l-2p, p) \in I_p(p,p).$$

Then $O(\iota_p) = (l-2p, 2p, l-2p, 2p)$, which is of type II. We will show that there are no other words of type II in J. Since we have m different values of O, Lemma 3.5 will imply (23) in the case w is of type II.

Let $K = J \setminus \{\alpha_p^{(p,p)}(\iota_p) \mid 0 \le p \le m-1\}$. We will find a partition of K into two sets, B and C, both with cardinality $2m(2m-1)(2m+1)/3$, and a bijection $\beta : B \to C$ such that $O(\beta(\iota)) = O(\iota)$ and check that O restricted to B is injective and its values are of type I. From this it will follow that (23) holds in the case w is of type I, and this will also complete the proof of (23) in the case w is of type II.

The partition and bijection are defined below in several parts. In all cases, it is straightforward to check the identity $O(\beta(i)) = O(i)$.

(i) For $0 \le p \le m-1$ take

$$B_1(p) = I_{p+1}(p,p),$$
$$C_1(p) = \{(s, l-p-s, 2p, l-p-t, t) \in I_p(p,p) \mid p+1 \le s, t\}.$$

We notice $B_1(p) = C_1(p)$ for all $0 \le p \le m-1$. This identification is used to define the restriction of β to $J_{p+1}(p,p)$ by $\beta \circ \alpha_{p+1}^{(p,p)} = \alpha_p^{(p,p)}$. For $\iota = (s, l-p-s, 2p, l-$

$p-t,t) \in I_{p+1}(p,p)$ we have

$$O(\iota)=\begin{cases}(l-p-s,2p,l-p-t,s+t), & p+1\le t\le s\le l-p,\\ (2p,l-p-t,s+t,l-p-s), & p+1\le s<t\le l-p,\end{cases}$$

and this element is of type I.

Let $B_1=\bigcup_{p=0}^{m-1}\alpha_{p+1}^{(p,p)}(B_1(p))$ and $C_1=\bigcup_{p=0}^{m-1}\alpha_p^{(p,p)}(C_1(p))$. We have

$$\#B_1=\sum_{p=0}^{m-1}(2(m-p)-1)^2.$$

(ii) For $1\le p\le m-1$, let

$$B_2(p)=\{(s,l-(p-1)-s,2p-1,l-p-t,t)\in I_p(p-1,p)\mid p+1\le s\},$$
$$C_2(p)=\{(\tilde{s},l-p-\tilde{s},2p-1,l-(p-1)-\tilde{t},\tilde{t})\in I_p(p,p-1)\mid p+1\le \tilde{t}\}.$$

For $\iota=(s,l-(p-1)-s,2p-1,l-p-t,t)\in B_2(p)$ let

$$\beta(\alpha_p^{(p-1,p)}(\iota))=\alpha_p^{(p,p-1)}(s-1,l-p-(s-1),2p-1,l-(p-1)-(t+1),t+1).$$

Then $\beta:\alpha_p^{(p-1,p)}(B_2(p))\to\alpha_p^{(p,p-1)}(C_2(p))$ is a bijection and a computation shows

$$\begin{aligned}&O(\beta(\alpha_p^{(p-1,p)}(\iota)))\\ &\quad=O(\alpha_p^{(p-1,p)}(\iota))\\ &\quad=\begin{cases}(l-p-s-1,2p-1,l-p-t,s+t), & p\le t\le s-1\le l-p,\\ (2p-1,l-p-t,s+t,l-p-s-1), & p\le s-1<t\le l-p\end{cases}\end{aligned}$$

and this is a word of type I. Take

$$B_2=\bigcup_{p=1}^{m-1}\alpha_p^{(p-1,p)}(B_2(p)),\qquad C_2=\bigcup_{p=1}^{m-1}\alpha_p^{(p,p-1)}(C_2(p)).$$

By disjointness, we have

$$\#B_2=\sum_{p=1}^{m-1}(2(m-p))^2.$$

(iii) In $I_0(0,0)$, the cases $(s,t)=(0,l)$ and $(s,t)=(l,0)$ have the same value under O, namely $(0,0,l,l)$, which is type I. Take

$$B_3=\{\alpha_0^{(0,0)}(l,0,0,l,0)\},\qquad C_3=\{\alpha_0^{(0,0)}(0,l,0,0,l)\}$$

and let $\beta(\alpha_0^{(0,0)}(l,0,0,l,0))=\alpha_0^{(0,0)}(0,l,0,0,l)$.

(iv) Consider the set

$$B_4(0)=\{(0,l,0,l-t,t):1\le t\le l-1\}\subset I_0(0,0).$$

For $\iota = (0, l, 0, l - t, t) \in B_4(0)$, take

$$\beta(\alpha_0^{(0,0)}(\iota)) = \begin{cases} \alpha_q^{(q,q)}(l - q, 0, 2q, l - 2q, q), & l - t \text{ even}, q = \frac{l-t}{2}, \\ \alpha_q^{(q,q-1)}(l - q, 0, 2q - 1, l - 2q + 1, q), & l - t \text{ odd}, q = \frac{l-t+1}{2}. \end{cases}$$

Let $B_4 = \alpha_0^{(0,0)}(B_4(0))$ and let C_4 be the image of B_4 under β. A direct computation shows

$$O(\beta(\alpha_0^{(0,0)}(\iota))) = O(\alpha_0^{(0,0)}(\iota)) = (0, l - t, t, l),$$

which is type I. We also have $\#B_4 = 2(m - 1)$.

(v) Consider the set

$$B_5(0) = \{(s, l - s, 0, l, 0) : 1 \le s \le l - 1\} \subset I_0(0, 0).$$

For $\iota = (s, l - s, 0, l, 0) \in B_5(0)$ define

$$\beta(\alpha_0^{(0,0)}(\iota)) = \begin{cases} \alpha_q^{(q,q)}(q, l - q, 2q, 0, l - q), & l - s \text{ even}, q = \frac{l-s}{2}, \\ \alpha_q^{(q-1,q)}(q, l - 2q + 1, 2q - 1, 0, l - q), & l - s \text{ odd}, q = \frac{l-s+1}{2}. \end{cases}$$

Let C_5 be the image of B_5 under β. Then $\beta : B_5 \to C_5$ is a bijection and

$$O(\beta(\alpha_0^{(0,0)}(\iota))) = O(\alpha_0^{(0,0)}(\iota)) = (l - s, 0, l, s)$$

is of type I. We also have $\#B_5 = 2(m - 1)$.

(vi) Let

$$B_6^1 = \bigcup_{p=1}^{m-1} \{\alpha_p^{(p-1,p)}(p, l - 2p + 1, 2p - 1, l - p - t, t) : p \le t \le l - p - 1\},$$

$$B_6^2 = \bigcup_{p=1}^{m-2} \{\alpha_p^{(p,p)}(p, l - 2p, 2p, l - p - t, t) : p + 1 \le t \le l - p - 1\}$$

and let $B_6 = B_6^1 \cup B_6^2$. For

$$\eta = \alpha_p^{(p-1,p)}(p, l - 2p + 1, 2p - 1, l - p - t, t) \in B_6^1, \tag{24}$$

let

$$\beta(\eta) = \begin{cases} \alpha_q^{(q,q)}(2m - 2p - q, 2p - 1, 2q, l - 2q, q), & p + t \text{ odd}, \\ \alpha_q^{(q,q-1)}(2m - 2p - q, 2p - 1, 2q - 1, l - 2q + 1, q), & p + t \text{ even}, \end{cases}$$

where $q = m - \lfloor \frac{p+t+1}{2} \rfloor$. For

$$\eta = \alpha_p^{(p,p)}(p, l - 2p, 2p, l - p - t, t) \in B_6^2 \tag{25}$$

let

$$\beta(\eta) = \begin{cases} \alpha_q^{(q,q)}(2m - 2p - q - 1, 2p, 2q, l - 2q, q), & p + t \text{ odd}, \\ \alpha_q^{(q,q-1)}(2m - 2p - q - 1, 2p, 2q - 1, l - 2q + 1, q) & p + t \text{ even}, \end{cases}$$

where $q = m - \lfloor \frac{p+t+1}{2} \rfloor$. Take C_6 to be the image of B_6 under β. Then $\beta : B_6 \to C_6$ is a bijection and

$$O(\beta(\eta)) = O(\eta) = \begin{cases} (2p-1, l-p-t+1, p+t, l-2p), & \eta \text{ as in (24)}, \\ (2p, lp-t, p+t, l-2p), & \eta \text{ as in (25)} \end{cases}$$

is of type I. We also have

$$\#B_6 = \sum_{p=1}^{m-1}(2(m-p)-1) + \sum_{p=1}^{m-2}(2(m-p)-2) = (2m-3)(m-1).$$

Lastly, we take

$$B = \bigsqcup_{k=1}^{6} B_k, \qquad C = \bigsqcup_{k=1}^{6} C_k.$$

A computation shows

$$\begin{aligned} \#B &= \sum_{p=1}^{m-1}(2(m-p)-1)^2 + \sum_{p=1}^{m-1}(2(m-p))^2 + 1 + 4(m-1) + (2m-3)(m-1) \\ &= 1 + (2m-1)^2 + 4(m-1) + (2m-3)(m-1) + \sum_{j=1}^{2(m-1)} j^2 \\ &= \frac{2m(2m-1)(2m+1)}{3}. \end{aligned}$$

We have, thus, constructed a bijection $\beta : B \to C$ that satisfies $O(\beta(\eta)) = O(\eta)$ and, as can be checked, the restriction of O to B is injective and takes values that are all of type I. Lastly the sets B and C form a partition of K. This completes the proof of Proposition 3.3.

The bijection we have defined may be better understood using some pictures, which are contained in Figs. 1 and 2. We parameterize I_0 by the square $\{(s,t) \in \mathbb{Z}^2 : 0 \le s, t \le l\}$ and I_m by the single point (m, m). Likewise for fixed $1 \le p \le m-1$ and $i, j \in \{p-1, p\}$, the set $I_p(i, j)$ is parameterized by $\{(s,t) \in \mathbb{Z}^2 : p \le s \le l-i,\ p \le t \le l-i\}$. We show the case $m = 3$.

In these figures,

- The points that give words of type II are marked with diamonds.
- The light circles in the right column are matched with the circles in the left. Likewise the solid circles. These correspond to cases 1 and 2.

  In the case 2 the bijection is implemented by $(s,t) \mapsto (s-1, t+1)$, form the rightmost sub-square of side $l-2p+1$ in $I_p(p-1, p)$ to the uppermost sub-square of side $l-2p+1$ in $I_p(p, p-1)$, for $1 \le p \le m-1$.
- Case 3 is marked with a solid square.
- The remaining points (which correspond to the most complicated part of the bijection), plotted in light squares, correspond the cases 4, 5 and 6. □

The following theorem summarizes the results obtained so far in this section.

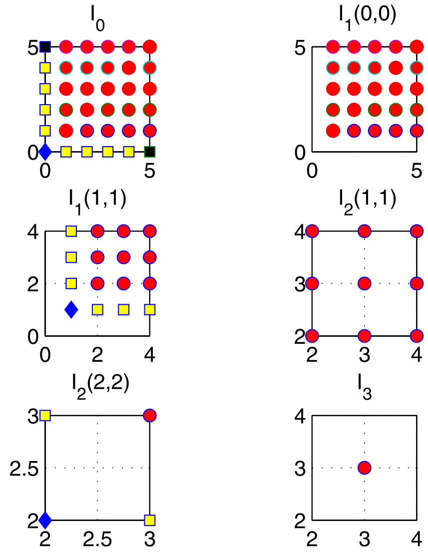

Fig. 1 Some sets in K with $m = 3$

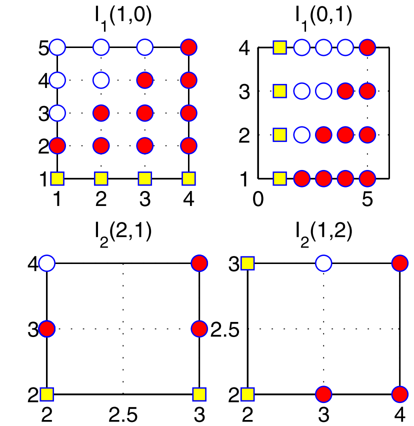

Fig. 2 More sets in K with $m = 3$

Theorem 3.7 *If $k = 2$ and $m \geq 2$ is even, or if $k = 4$ and $m \geq 6$ is even but not a multiple of 4, then $S_{m,k}(A, B)$ is cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle A, B\rangle$. Therefore, for these values of m and k, $\mathrm{Tr}(S_{m,k}(A, B)) \geq 0$ whenever A and B are Hermitian matrices.*

Below is a non-sum-of-squares result for $S_{8,4}(A, B)$. However, Question 1.1 for $m = 8$ and $k = 4$ is still open.

Proposition 3.8 *The polynomial $S_{8,4}(A, B)$ is not cyclically equivalent to a sum of Hermitian squares in $\mathbf{R}\langle A, B\rangle$.*

Proof We order the elements of $W_{2,2}(A, B)$ in the column vector

$$Z = (A^2B^2, ABAB, AB^2A, BA^2B, BABA, B^2A^2)^t.$$

Table 9 Representatives of cyclic equivalence classes in $W_{4,4}(A, B)$.

name	word	order
w_1	A^4B^4	8
w_2	A^3BAB^3	8
w_3	$A^3B^2AB^2$	8
w_4	A^3B^3AB	8
w_5	$A^2BA^2B^3$	8
w_6	A^2BABAB^2	8
w_7	A^2BAB^2AB	8
w_8	$A^2B^2A^2B^2$	4
w_9	A^2B^2ABAB	8
w_{10}	$ABABABAB$	2

If $S_{8,4}(A, B)$ were equivalent to a sums of squares in $\mathbf{R}\langle A, B\rangle$, then by Proposition 3.3 of [6], we would have $S_{8,4}(A, B) \overset{\text{cyc}}{\sim} Z^*HZ$ for H a 6×6 real, positive semidefinite matrix. So suppose, to obtain a contradiction, that such exists. There are ten cyclic equivalence class of words in $W_{4,4}(A, B)$. We've chosen one representative for each and we have listed them in Table 9 with their orders, where we say the *order* of a word is the number of cyclically equivalent forms that it has. If we denote the ith element of the vector Z by z_i, then the matrix whose (i, j)th entry is the symbol $k \in \{1, \dots, 10\}$ such that w_k is cyclically equivalent to $z_i^* z_j$ is the matrix found below.

$$\begin{pmatrix} 1 & 2 & 3 & 5 & 6 & 8 \\ 4 & 7 & 9 & 9 & 10 & 6 \\ 3 & 6 & 8 & 7 & 9 & 3 \\ 5 & 6 & 7 & 8 & 9 & 5 \\ 9 & 10 & 6 & 6 & 7 & 2 \\ 8 & 9 & 3 & 5 & 4 & 1 \end{pmatrix}.$$

The hypothesis $Z^*HZ \sim S_{8,4}(A, B)$ is, therefore, equivalent to the ten linear equations

$$8 = H_{11} + H_{66}, \tag{26}$$

$$8 = H_{12} + H_{56}, \tag{27}$$

$$8 = H_{13} + H_{31} + H_{36} + H_{63}, \tag{28}$$

$$8 = H_{21} + H_{65}, \tag{29}$$

$$8 = H_{14} + H_{41} + H_{46} + H_{64}, \tag{30}$$

$$8 = H_{15} + H_{26} + H_{32} + H_{42} + H_{53} + H_{54}, \tag{31}$$

$$8 = H_{22} + H_{34} + H_{43} + H_{55}, \tag{32}$$

$$4 = H_{16} + H_{33} + H_{44} + H_{61}, \tag{33}$$

$$8 = H_{23} + H_{24} + H_{35} + H_{45} + H_{51} + H_{62}, \tag{34}$$

$$2 = H_{25} + H_{52} \tag{35}$$

in the entries of the matrix H. However, H is real symmetric. Moreover, we may assume without loss of generality that the relations (4) and (7) from Remark 2.1 hold, and we find, therefore, that H commutes with the permutation matrices corresponding to the order-two permutations

$$\tau : 1 \leftrightarrow 6,\ 2 \leftrightarrow 5.$$

$$\sigma : 1 \leftrightarrow 6,\ 2 \leftrightarrow 5,\ 3 \leftrightarrow 4.$$

Thus, we have

$$H = \begin{pmatrix} H_{11} & H_{12} & H_{13} & H_{13} & H_{15} & H_{16} \\ H_{12} & H_{22} & H_{23} & H_{23} & H_{25} & H_{15} \\ H_{13} & H_{23} & H_{33} & H_{34} & H_{23} & H_{13} \\ H_{13} & H_{23} & H_{34} & H_{33} & H_{23} & H_{13} \\ H_{15} & H_{25} & H_{23} & H_{23} & H_{22} & H_{12} \\ H_{16} & H_{15} & H_{13} & H_{13} & H_{12} & H_{11} \end{pmatrix}.$$

Equations (26)–(35) now yield several relations, for example, from (26) we get $H_{11} = 4$. Using these relations to eliminate some variables, we have that H equals the matrix

$$\begin{pmatrix} 4 & 4 & 2 & 2 & 4-2H_{23} & 2-H_{33} \\ 4 & H_{22} & H_{23} & H_{23} & 1 & 4-2H_{23} \\ 2 & H_{23} & H_{33} & 4-H_{22} & H_{23} & 2 \\ 2 & H_{23} & 4-H_{22} & H_{33} & H_{23} & 2 \\ 4-2H_{23} & 1 & H_{23} & H_{23} & H_{22} & 4 \\ 2-H_{33} & 4-2H_{23} & 2 & 2 & 4 & 4 \end{pmatrix}.$$

We will show that there is no positive semidefinite real matrix of this form. To make the formulas slightly more readable, we will use the symbols $x_2 = H_{22}$ and $x_3 = H_{33}$. Of course, we must have $x_2 \geq 0$ and $x_3 \geq 0$. We will consider compressions of H obtained by restricting to rows and columns in subsets of $\{1, \dots, 6\}$. The compression to $\{1, 2\}$ is $\left(\begin{smallmatrix} 4 & 4 \\ 4 & x_2 \end{smallmatrix}\right)$, and from positivity we obtain $x_2 \geq 4$. Compression to $\{1, 6\}$ yields $|2 - x_3| \leq 4$, so $x_3 \leq 6$. Compression to $\{1, 3\}$ yields $x_3 \geq 1$. The determinant of the compression of the matrix H to $\{1, 3, 4, 6\}$ is the polynomial with factorization

$$(2 + x_3)(x_2 + x_3 - 4)(8 - 6x_2 + 2x_3 + x_2x_3 - x_3^2).$$

Since $x_3 \geq 1$ and $x_2 \geq 4$, the first two factors are strictly positive. So the third factor must be nonnegative, and we conclude

$$x_2(x_3 - 6) \geq (x_3 - 4)(x_3 + 2).$$

Since $x_3 \leq 6$ we must have $x_3 \leq 4$ and

$$x_2 \leq \frac{(4 - x_3)(x_3 + 2)}{6 - x_3}.$$

But combining this with $x_2 \geq 4$, we get $24 - 4x_3 \leq 8 + 2x_3 - x_3^2$, so $x_3^2 - 6x_3 + 16 \leq 0$, which is impossible. This is the desired contradiction. □

4 Proof of Theorem 1.2

In this section, we prove Theorem 1.2 using a straightforward application of the method of Lagrange multipliers.

Lemma 4.1 *Let $n, m, k \in \mathbf{N}$ and fix an $n \times n$ Hermitian matrix B. Consider the function $A \mapsto \mathrm{Tr}(S_{m,k}(A, B))$ with domain consisting of the $n \times n$ Hermitian matrices A such that $\mathrm{Tr}(A^2) = 1$. Suppose A_0 is a point where this function has a relative extremum. Then*

$$S_{m-1,k}(A_0, B) = \frac{m-k}{m} \mathrm{Tr}(S_{m,k}(A_0, B))A_0. \tag{36}$$

Proof This is an application of the method of Lagrange multipliers to the problem of optimizing $\mathrm{Tr}(S_{m,k}(A, B))$ subject to the constraint $\mathrm{Tr}(A^2) = 1$. (Compare to Appendix A of [6].) The space of Hermitian $n \times n$ matrices is a real vector space of dimension n^2. If H and A are Hermitian matrices, then

$$\left.\frac{d}{d\lambda}\right|_{\lambda=0} \mathrm{Tr}((A + \lambda H)^2) = 2\,\mathrm{Tr}(HA). \tag{37}$$

Letting H run through a fixed basis for the space of $n \times n$ Hermitian matrices, the list of values (37) forms the gradient of the constraint function with respect to the n^2 variables.

Letting $W_{m-k,k}(A, B)$ be the set of all words in noncommuting variables A and B with $m-k$ A's and k B's, we have $|W_{m-k,k}(A, B)| = \binom{m}{k}$. If $w = w(A, B) \in W_{m-k,k}(A, B)$, then $\frac{d}{d\lambda}\big|_{\lambda=0} w(A + \lambda H, B)$ equals the sum of the $m-k$ words obtained by replacing in turn and individually the letters of w that are equal to A by H. Thus, $\frac{d}{d\lambda}\big|_{\lambda=0} S_{m,k}(A + \lambda H, B)$ is the sum of all $(m-k)\binom{m}{k}$ words in A, B and H, where A appears $m-k-1$ times, B appears k times and H appears once. Taking the trace, we get

$$\left.\frac{d}{d\lambda}\right|_{\lambda=0} \mathrm{Tr}(S_{m,k}(A + \lambda H, B)) = m\,\mathrm{Tr}(HS_{m-1,k}(A, B)). \tag{38}$$

Letting H run through the same basis as taken above, the list of values (38) forms the gradient of the objective function with respect to the n^2 variables.

By the method of Lagrange multipliers, we conclude that at a relative extremum A_0, these two gradients must be parallel. In other words, we must have

$$2\mu\,\mathrm{Tr}(HA_0) = m\,\mathrm{Tr}(HS_{m-1,k}(A_0, B))$$

for some $\mu \in \mathbf{R}$ and all H, and this implies

$$2\mu A_0 = mS_{m-1,k}(A_0, B).$$

Multiplying both sides by A_0, taking the trace and using Lemma 2.1 of [4], we get

$$2\mu = 2\mu\,\mathrm{Tr}(A_0^2) = m\,\mathrm{Tr}(A_0 S_{m-1,k}(A_0, B)) = (m-k)\,\mathrm{Tr}(S_{m,k}(A_0, B)),$$

and (36) follows. □

Proof of Theorem 1.2 The implication (1) $\Longrightarrow$ (2) is clear.

Suppose (1) does not hold. Let A_0 and B_0 be a Hermitian $n \times n$ matrices where $\mathrm{Tr}(S_{m,k}(A, B))$ takes its absolute minimum subject to $\mathrm{Tr}(A^2) = \mathrm{Tr}(B^2) = 1$. By assumption, we have $\mathrm{Tr}(S_{m,k}(A_0, B_0)) < 0$. By Lemma 4.1 and the analogue obtained by switching A and B, we have

$$S_{m-1,k}(A_0, B_0) = \frac{m-k}{m} \mathrm{Tr}(S_{m,k}(A_0, B_0))A_0,$$

$$S_{m,k-1}(A_0, B_0) = \frac{k}{m} \operatorname{Tr}(S_{m,k}(A_0, B_0)) B_0.$$

Therefore, the Hermitian matrix

$$\begin{aligned} S_{m,k}(A_0, B_0) &= A_0 S_{m-1,k}(A_0, B_0) + B_0 S_{m-1,k-1}(A_0, B_0) \\ &= \operatorname{Tr}(S_{m,k}(A_0, B_0)) \left(\frac{m-k}{m} A_0^2 + \frac{k}{m} B_0^2 \right) \end{aligned}$$

has only nonpositive eigenvalues. Thus, (2) does not hold. □

Acknowledgement The authors thank an anonymous referee for suggestions that improved the exposition.

References

1. Bessis, D., Moussa, P., Villani, M.: Monotonic converging variational approximations to the functional integrals in quantum statistical mechanics. J. Math. Phys. **16**, 2318–2325 (1975)
2. Burgdorf, S.: Sums of Hermitian squares as an approach to the BMV conjecture. Linear Multilinear Algebra (to appear)
3. Hägele, D.: Proof of the cases $p \leq 7$ of the Lieb–Seiringer formulation of the Besis–Moussa–Villani conjecture. J. Stat. Phys. **127**, 1167–1171 (2007)
4. Hillar, C.J.: Advances on the Bessis–Moussa–Villani trace conjecture. Linear Algebra Appl. **426**, 130–142 (2007)
5. Hillar, C.J., Johnson, C.R.: On the positivity of the coefficients of a certain polynomial defined by two positive definite matrices. J. Stat. Phys. **118**, 781–789 (2005)
6. Klep, I., Schweighofer, M.: Sums of Hermitian squares and the BMV conjecture. J. Stat. Phys. **133**, 739–760 (2008)
7. Klep, I., Schweighofer, M.: Connes' embedding conjecture and sums of Hermitian squares. Adv. Math. **217**, 1816–1837 (2008)
8. Landweber, P.S., Speer, E.R.: On D. Hägele's approach to the Bessis–Moussa–Villani conjecture. Linear Algebra Appl. **431**, 1317–1324 (2009)
9. Lieb, E.H., Seiringer, R.: Equivalent forms of the Bessis–Moussa–Villani conjecture. J. Stat. Phys. **115**, 185–190 (2004)
10. Wolfram Research, Inc.: Mathematica Version 7.0. Wolfram Research, Inc., Champaign (2008)