TRIDIAGONAL FORMS IN LOW DIMENSIONS

KENNETH R. DAVIDSON AND DRAGOMIR Ž. ĐOKOVIĆ

Abstract. Pati showed that every 4×4 matrix is unitarily similar to a tridiagonal matrix. We give a simple proof. In addition, we show that (in an appropriate sense) there are generically precisely 12 ways to do this. When the real part is diagonal, it is shown that the unitary can be chosen with the form $U = PD$ where D is diagonal and P is real orthogonal. However even if both real and imaginary parts are real symmetric, there may be no real orthogonal matrices which tridiagonalize it. On the other hand, if the matrix belongs to the Lie algebra $\mathfrak{sp}_4(\mathbb{C})$, then it can be tridiagonalized by a unitary in the symplectic group Sp(2). In dimension 5 or greater, there are always rank three matrices which are not tridiagonalizable.

A matrix T can be tridiagonalized if there is an orthonormal basis $e_1, e_2, \dots, e_n$ so that $\langle Te_j, e_i\rangle = 0$ for $|i-j| > 1$. Most matrices do not have this restricted form. Indeed, Longstaff [6] and Fong and Wu [3] showed that if $n \geq 5$, there are $n\times n$ matrices which do not have this form. On the other hand, it is easy to see that any 3×3 matrix and any rank two matrix is tridiagonalizable. The case of 4×4 matrices proved to be very difficult. This was solved by Pati [7] who used methods of algebraic geometry to establish that every 4×4 matrix is tridiagonalizable. The real case was investigated by the second author and MacDonald [2] where they produced a real 4×4 matrix which cannot be tridiagonalized by a real orthogonal matrix.

The main purpose of this paper is to provide a simpler proof of a sharper version of Pati's Theorem. The sharpening arises because tridiagonalization is reduced to solving a simple set of polynomial equations which are homogeneous in two sets of variables. Bezout's Theorem allows us to count precisely the number of solutions. The precise details of the genericity argument take us away from the main thrust, which is that tridiagonalization is always possible for 4×4 matrices. So we leave that argument to a separate section. The reader who is only interested in existence can safely skip it.

The authors are partially supported by NSERC.

2000 *Mathematics Subject Classification.* Primary 15A21; Secondary 15A22.

We investigate certain consequences of Pati's Theorem. Tridiagonalization of T is equivalent to simultaneously tridiagonalizing two Hermitian matrices—the real and imaginary part of the original, namely $A = (T + T^*)/2$ and $B = (T - T^*)/2i$. If the real part is first diagonalized, then we show that the unitary has the special form $U = PD$ where D is diagonal and P is real orthogonal. This raised the natural question of whether, when both Hermitian matrices are real symmetric, the unitary can be chosen to be real orthogonal. This is not the case. An example was computer generated using Maple by explicitly finding all 12 possible tridiagonalizations.

This problem can be considered in the context of tridiagonalizing elements of a Lie algebra by unitaries in the maximal compact subgroup of the corresponding Lie group. In most cases, this is impossible by dimension arguments—just as in the classical case. In the small number of cases not eliminated in this way, there is one other interesting case. This may be described as tridiagonalizing elements of the Lie algebra $\mathfrak{sp}_4(\mathbb{C})$ by a symplectic similarity. We show that this is always possible.

We also provide a sharpening of the results of Longstaff [6] and Fong and Wu [3] in higher dimensions and Hilbert space. We show that in dimension $n \geq 5$ including infinite dimensional Hilbert space, there are rank 3 matrices which cannot be tridiagonalized. In addition to a soft dimension argument, we also provide an explicit example.

1. Tridiagonalizable Matrices

A matrix T acting on n-dimensional complex inner product space $\mathcal{H}_n = \mathbb{C}^n$ is *tridiagonal* with respect to an orthonormal basis $e_1, \ldots, e_n$ if $\langle Te_j, e_i \rangle = 0$ if $|i - j| > 1$. Say that T is *tridiagonalizable* if such a basis exists. This is equivalent to saying that there is a unitary matrix U so that UTU^* is tridiagonal. Similarly an operator acting on a separable Hilbert space $\mathcal{H}$ has a *tridiagonal form* if it can be written as an orthogonal direct sum $T = \sum \oplus T_k$ where each T_k has a tridiagonal form in which the basis for each T_k is indexed by an interval of $\mathbb{Z}$.

Since we will be working exclusively with unitary similarities, we will use the term basis to refer to an orthonormal basis.

If T has a tridiagonal form, then so does the whole subspace spanned by $\{I, T, T^*\}$. This is also spanned by $\{I, A, B\}$ where A and B are the real and imaginary parts of T. So the problem of tridiagonalizing a matrix T is equivalent to the problem of simultaneously tridiagonalizing the pair $\{A, B\}$ of Hermitian matrices.

The set of $n \times n$ matrices which are tridiagonalized by a given basis is a closed subspace $\mathcal{T}$ of $\mathcal{M}_n$. The set of tridiagonalizable matrices is the image of the map from $\mathrm{U}(n) \times \mathcal{T}$ to $\mathcal{M}_n$ given by $\tau(U,T) = UTU^*$. Since $\mathrm{U}(n)$ is compact, this image is closed. Hence the set of tridiagonalizable matrices is closed; and the set of non-tridiagonalizable matrices is open. So it suffices to tridiagonalize a generic set of matrices in order to establish that all are tridiagonalizable.

If T is tridiagonal with respect to the basis $e_1, e_2, \dots, e_n$, define subspaces $E_j = \operatorname{span}\{e_i : i \le j\}$. Observe that this is a *flag*, i.e. a nested sequence of subspaces with $\dim(E_{j+1}/E_j) = 1$ for $1 \le j < n$, such that

$$E_j + TE_j + T^*E_j = E_j + AE_j + BE_j \subset E_{j+1}.$$

Conversely, if E_j for $1 \le j \le n$ is such a flag, choose any orthonormal basis so that $e_j \in E_j \ominus E_{j-1}$. It follows that $\langle Te_j, e_i\rangle = 0$ if $i \ge j+2$ and likewise $\langle T^*e_j, e_i\rangle = 0$ if $i \ge j+2$. Therefore $\langle Te_j, e_i\rangle = 0$ if $|i-j| \ge 2$, meaning that T is tridiagonal.

For example, this leads to the easy solution in dimension 3. Given T, select any eigenvector e and take $V_1 = \mathbb{C}e$ and $V_2 = \operatorname{span}\{e, T^*e\}$. When e is also an eigenvector for T^*, any orthonormal basis beginning with e will work. Fong and Wu [3] also establish the easy result that every matrix of rank 2 can be tridiagonalized. In dimensions $n \ge 5$, Longstaff [6] and Fong and Wu establish that there are matrices which cannot be tridiagonalized by using a dimension argument.

In dimension four, V. Pati [7] proved the following result. It is interesting because the proof is a very difficult argument in algebraic geometry. A new proof will be provided in the next section. It explains why such methods are needed, because in general the number of solutions is very small.

Theorem 1.1 (Pati). *Every complex 4×4 matrix is unitarily similar to a tridiagonal matrix. Equivalently, if A and B are 4×4 Hermitian matrices, then there exists a unitary matrix U such that UAU^* and UBU^* are tridiagonal.*

In the remainder of this section, we establish a few results that will finish off a direction begun in [6, 3] about dimensions other than 4.

Proposition 1.2. *Suppose that T is a finite rank matrix or operator on Hilbert space which has a tridiagonal form. Then it has a tridiagonal form when restricted to $M = \operatorname{Ran} T \vee \operatorname{Ran} T^*$.*

Proof. We may suppose that T is irreducible on M in the sense that T and T^* have no common invariant subspaces in M. Otherwise, decompose T into an orthogonal direct sum of irreducible summands and treat each term separately. Also an elementary observation [3] is that the tridiagonal form of a rank d matrix must be supported on at most

$3d$ of the tridiagonalizing basis vectors. So we may suppose that T is a matrix on an m-dimensional space.

Let $M = \operatorname{Ran} T \vee \operatorname{Ran} T^*$. Note that $M^\perp = \ker T \cap \ker T^*$. So T decomposes as $T \simeq T|_M \oplus 0_{M^\perp}$. Let P_M denote the orthogonal projection onto M.

Let $e_1, e_2, \dots, e_m$ be an orthonormal basis tridiagonalizing T. We may suppose that either the first row or column of the matrix is non-zero. Let $E_k = \operatorname{span}\{e_1, \dots, e_k\}$ be the associated flag. If the containment $E_j + TE_j + T^*E_j \subset E_{j+1}$ is proper, then $TE_k + T^*E_k \subset E_k$; so E_k is a reducing subspace of T. By the irreducibility hypothesis, this means that $E_k \supset M$ and hence the tridiagonalization is complete (except for a zero summand).

Now replace E_k by $V_k := P_M E_k$ for $k \geq 1$. Observe that since T and T^* commute with P_M,

$$\begin{aligned} V_k + TV_k + T^*V_k &= P_M(E_k + TE_k + T^*E_k) \\ &\subset P_M E_{k+1} = V_{k+1}. \end{aligned}$$

By hypothesis $V_1 \neq \{0\}$. Clearly $\dim(V_{k+1}/V_k) \leq \dim(E_{k+1}/E_k) = 1$. Moreover, if $V_{k+1} = V_k$, then V_k reduces T; whence $V_k = M$ as above. Hence this occurs exactly when $k = d$; i.e. $V_d = M$.

Choose an orthonormal basis $v_1, \dots, v_d$ for M by selecting a unit vector v_i in $V_i \ominus V_{i-1}$ for $1 \leq i \leq d$. Evidently this is the desired tridiagonalizing basis. ∎

Corollary 1.3. *Every tridiagonalizable Hilbert space operator of rank d can be tridiagonalized with a basis $e_1, \dots, e_{2d}$ direct summed with an infinite zero operator.*

The following result is immediate from results in the literature. We strengthen it in Proposition 1.6 below.

Corollary 1.4. *There is an operator T of rank 4 on Hilbert space which cannot be tridiagonalized.*

Proof. By [3, Theorem 2.1], there is a 5×5 matrix T which cannot be tridiagonalized. This property is not affected by adding a scalar. So we may suppose that T is singular, and thus has rank at most 4. By Proposition 1.2, the operator $T \oplus 0_\infty$ cannot be tridiagonalized either. (Here 0_∞ is a zero operator of infinite multiplicity.)

If T has rank less than 4 (namely 3), then one can use the fact that the set of non-tridiagonalizable operators is an open set to find a rank four matrix sufficiently near to T. ∎

Corollary 1.5. *If T is a tridiagonal operator which decomposes as an orthogonal direct sum $T = T_1 \oplus T_2$, then both T_1 and T_2 can be tridiagonalized.*

Proof. Let P be the orthogonal projection onto the domain of the first summand T_1. As in the proof of Proposition 1.2, the subspaces $V_j = PE_j$ form a tridiagonalizing flag for T_1. ∎

We show that similar dimension arguments also apply to rank 3 matrices in dimension at least 5. The previous analysis together with Pati's Theorem shows that if $\dim(\operatorname{Ran} T \vee \operatorname{Ran} T^*) \leq 4$, then T can be tridiagonalized. For rank 3 matrices, this dimension can be 5 or 6. In section 7 we will provide an explicit example.

Proposition 1.6. *There exist rank three 5×5 matrices which are not tridiagonalizable.*

Proof. Let $\mathcal{T}$ denote the subspace of $\mathcal{M}_5$ consisting of all tridiagonal matrices, and let $\widetilde{\mathcal{T}}$ denote the set of all matrices in $\mathcal{M}_5$ which are (unitarily) tridiagonalizable. Then $\widetilde{\mathcal{T}} = \mathrm{U}(5) \cdot \mathcal{T}$, where the dot denotes the conjugation action $(X, A) \to XAX^{-1}$ of $\mathrm{U}(5)$ on $\mathcal{M}_5$. However it is clear that subgroup T^5 of diagonal unitaries conjugates $\mathcal{T}$ onto itself. Thus it suffices to use the set W of unitary matrices with real first column. So $\widetilde{\mathcal{T}} = \mathrm{W} \cdot \mathcal{T}$. Recall that $\mathrm{U}(5)$ has real dimension 25. Hence W has real dimension $25 - 5 = 20$.

Let $\mathcal{D}_3$ denote the set of 5×5 matrices of rank at most 3. In general, the determinantal variety of $m \times n$ complex matrices of rank at most r is irreducible of complex dimension $(m + n - r)r$ (see [5, Prop. 12.2]). We are interested in $\mathcal{D}_3$, the case $m = n = 5$ and $r = 3$; so its complex dimension is 21. We shall work with real dimensions. So $\dim \mathcal{D}_3 = 42$.

We set $\widetilde{\mathcal{D}_3} := \mathcal{D}_3 \cap \widetilde{\mathcal{T}}$ and $\mathcal{D}_3\mathcal{T} := \mathcal{D}_3 \cap \mathcal{T}$. Note that $\widetilde{\mathcal{D}_3} = \mathrm{W} \cdot \mathcal{D}_3\mathcal{T}$. Clearly $\mathcal{T}$ has real dimension $\dim \mathcal{T} = 26$. However every tridiagonal matrix of rank 3 must have a zero entry on the superdiagonal, for otherwise the rank is at least 4. Similarly the subdiagonal must have a zero entry. Also the determinant is zero. Hence the components of the variety $\mathcal{D}_3\mathcal{T}$ must have dimension no greater than $26 - 6 = 20$. In fact, each component has dimension 18. See Remark 7.1.

Since $\dim \mathrm{W} + \dim \mathcal{D}_3\mathcal{T} \leq 40 < \dim \mathcal{D}_3$, it follows that $\widetilde{\mathcal{D}_3}$ is nowhere dense in $\mathcal{D}_3$. ∎

As in Corollary 1.4, we obtain an immediate consequence:

Corollary 1.7. *There is an operator T of rank 3 on Hilbert space which cannot be tridiagonalized.*

2. A Proof of Pati's Theorem

The purpose of this section is to provide a new, much simpler proof of Pati's Theorem. The key idea is to produce a simple set of bi-homogeneous equations (i.e. equations jointly homogeneous in two sets of variables) which solves the tridiagonalization problem. Then the issue is to compute the multiplicities of a few extraneous solutions.

Bezout's Theorem is all that is needed. Indeed, it shows immediately that there are generically at most 12 flags which tridiagonalize a given matrix. In the next section, we will provide a more detailed analysis to show that generically one obtains exactly 12 flags.

Observe that in the case of 4×4 matrices, the tridiagonal matrices have real dimension 20. The real variety of unitary matrices with first column in $\mathbb{R}^4$ is $3 + 5 + 3 + 1 = 12$. The sum is 32, precisely the real dimension of $\mathcal{M}_4(\mathbb{C})$. So the dimension argument suggests that although solutions may exist, that generically they are 0-dimensional and hence likely there is only a finite set.

Theorem 2.1. *Every complex* 4×4 *matrix is unitarily equivalent to a tridiagonal matrix. Equivalently, if* A *and* B *are* 4×4 *Hermitian matrices, then there exists a unitary matrix* U *such that* UAU^* *and* UBU^* *are tridiagonal. Generically, there are at most* 12 *flags which tridiagonalize a given matrix.*

Proof. We deal with a pair A and B of Hermitian 4×4 matrices. We may assume that A is diagonal with respect to the orthonormal basis e_1, e_2, e_3, e_4. In addition, we make two generic assumptions:

(i) A has distinct eigenvalues $\alpha_1, \alpha_2, \alpha_3, \alpha_4$.
(ii) $\{e_i, Be_i, B^2e_i, ABe_i\}$ are linearly independent for $1 \le i \le 4$.

Indeed it is enough to solve the problem in this case because the set of tridiagonalizable matrices is closed. However it is also easy to see that a failure of condition (ii) leads immediately to a tridiagonalization. See Remark 2.3. In particular, (ii) guarantees that A and B have no common eigenvectors.

Suppose that A and B are tridiagonal with respect to a basis v_1, v_2, v_3, v_4. Let $s = \langle Bv_1, v_2 \rangle$ and $t = \langle Av_1, v_2 \rangle$. By the previous paragraph, s and t are not both 0. Observe that v_1 is an eigenvector for $sA - tB$.

Conversely if u is an eigenvector for some non-zero linear combination $sA + tB$, then

$$V_2 = \operatorname{span}\{u, Au, Bu\} = \operatorname{span}\{u, (sA + tB)u, (tA - sB)u\}$$

will always be two dimensional. Moreover, as long as u is not an eigenvector e_i of A, $V_2 = \text{span}\{u, Au\}$. Consider the subspace

$$V_3 = \text{span}\{u, Au, A^2u, BAu\} = V_2 + AV_2 + BV_2.$$

If V_3 is 3-dimensional, then the chain $V_1 \subset V_2 \subset V_3$ tridiagonalizes A and B. Thus we are looking for $\det(u, Au, A^2u, BAu) = 0$.

This leads us to the following system of equations in projective variables $[s : t : \lambda], [u_1 : u_2 : u_3 : u_4]$ in $\mathbb{C}P^2 \times \mathbb{C}P^3$.

$$(sA + tB - \lambda I)u = 0 \tag{1}$$

$$\det(u, Au, A^2u, BAu) = 0 \tag{2}$$

The first (1) is a set of four homogeneous equations of degree 1 in $[s : t : \lambda]$ and also degree 1 in $[u_1 : u_2 : u_3 : u_4]$. The second equation (2) is homogeneous of degree 4 in $[u_1 : u_2 : u_3 : u_4]$ and degree 0 in $[s : t : \lambda]$ as they do not appear.

Observe first that if $t = 0$, then we must have $s \neq 0$. So after normalization $s = 1$; and it follows easily that $\lambda = \alpha_i$ for some i. These are extraneous solutions that do not lead to a tridiagonalization. We claim that this system has a non-trivial solution $[s : t : \lambda] \neq 0$ and $u \neq 0$ such that $t \neq 0$. It is easy to see that any such solution gives a tridiagonalizing flag. If the vectors u, Au, A^2u are linearly dependent and u, Au, BAu independent, then one should take $V_3 = \text{span}\{u, Au, BAu\}$. If also u, Au, BAu are linearly dependent, then V_3 can be chosen arbitrarily.

If the system has infinitely many solutions, then clearly the above claim is valid. So assume that the system has only finitely many solutions. Then the Bezout Theorem is applicable.

By Bezout's Theorem for the product of two projective spaces (c.f. [8, IV.2]), either the number of solutions is infinite or it is a specific finite number. It is obtained by summing, over all possible partitions of the five equations into two equations and the remaining three, the product of the degrees in $[s : t : \lambda]$ for the first two and of u for the other three. Clearly this product is 0 unless the first two are two of the four bilinear equations. So there are $\binom{4}{2} = 6$ such choices which yields a total degree of $6(1 \cdot 1)(1 \cdot 1 \cdot 4) = 24$ solutions in the generic case.

The extraneous solutions mentioned above arise because we chose to define V_2 as $\text{span}\{u, Au\}$. This is not the right object in the exceptional case in which $u = e_i$ is an eigenvector of A. So in this case, there is a solution of our system, namely

$$s = 1,\ t = 0,\ \lambda = \alpha_i;\quad u_i = 1,\ u_j = 0 \text{ for } j \neq i$$

Moreover, you can convince yourself that this is the only way $u = e_i$ can arise as a solution. To count the number of bona fide solutions, we need to compute the multiplicity at each of these extraneous solutions.

The multiplicity is defined as the codimension of the ideal generated by the polynomials in our system localized at the solution point. One appraoch is to plug the system into the symbolic algebra program Singular and compute the Gröbner basis. This was done, and the multiplicity is 3. The single condition needed to ensure this is that

$$\det(e_i, Be_i, B^2e_i, ABe_i) \neq 0$$

which follows from our genericity assumption (ii). We derive this fact directly in Lemma 2.2 below. So the four extraneous solutions actually account for 12 solutions. That leaves exactly 12 other solutions.

Observe that a chain $V_1 \subset V_2 \subset V_3 \subset \mathbb{C}^4$ does not yield a unique unitary. Rather, it determines the subspaces V_1, $V_2 \ominus V_1$, $V_3 \ominus V_2$ and $V_3^\perp$ spanned by the tridiagonalizing basis. So the unitary is only unique up to multiplication by a diagonal unitary. However the flag is uniquely determined by the solution. Thus generically there are at most 12 flags which tridiagonalize A and B. Example 5.1 provides an example in which there are exactly 12 solutions, evidently all multiplicity one. ∎

Lemma 2.2. *The multiplicity of the system* (1) *and* (2) *at the extraneous solution* $([1:0:\alpha_i], u \in \mathbb{C}^* e_i)$ *is* 3 *provided that*

$$\det(e_i, Be_i, ABe_i, B^2e_i) \neq 0.$$

Proof. Let us compute the multiplicity for the extraneous solution $[1:0:\alpha_1]$, $[1:0:0:0]$. For definiteness, we set $s = 1$ and $u_1 = 1$. Let $\mathcal{J}$ be the ideal generated by our polynomials in the local ring over this point. The unique maximal ideal $\mathcal{I}$ in the local ring is generated by $\{\alpha_1 - \lambda, u_2, u_3, u_4, t\}$.

Write $B = [b_{ij}]$. The bilinear terms from $(A + tB - \lambda I)u$ can be written as

$$(\alpha_i - \lambda)u_i + t\sum_{j=1}^{4} b_{ij}u_j \quad \text{for} \quad 1 \leq i \leq 4.$$

Using $u_1 = 1$, observe that these four polynomials yield four generators in $\mathcal{J}$ which have leading term $\alpha_1 - \lambda$, u_2, u_3 and u_4 respectively. Let $\mathcal{J}_0$ be the ideal that they generate. Thus the problem becomes reducing equation (2) to yield a leading term which is a power of t.

Observe that $(A - \lambda I)u \equiv -tBu \pmod{\mathcal{J}_0}$. We repeatedly use this substitution in (2). So computing modulo $\mathcal{J}_0$ in the local ring,

$$\det(u, Au, A^2u, BAu) = \det\big(u, (A - \lambda I)u, (A^2 - \lambda^2 I)u, BAu\big)$$

$$\begin{aligned}
&\equiv \det\big(u, -tBu, -(A+\lambda I)tBu, BAu\big)\\
&= t^2 \det(u, Bu, ABu, BAu)\\
&= t^2 \det\big(u, Bu, ABu, B(A-\lambda I)u\big)\\
&\equiv t^2 \det(u, Bu, ABu, -tB^2u)\\
&= -t^3 \det(u, Bu, ABu, B^2u)
\end{aligned}$$

The order 3 part of this expression is $t^3 \det(e_1, Be_1, B^2e_1, ABe_1)$. This is a non-zero multiple of t^3 by our genericity assumption. Therefore the ideal has codimension exactly 3. ∎

Remark 2.3. Suppose that the generic condition (ii) fails, i.e.

$$V := \operatorname{span}\{e_i, Be_i, B^2e_i, ABe_i\}$$

is a proper subspace of $\mathbb{C}^4$. If $\dim V = 1$, e_i is a common eigenvector for A and B. This reduces the problem to the 3×3 case in $V^\perp$. If $\dim V = 2$, then V and $V^\perp$ are invariant for A and B. So any basis for V followed by a basis for $V^\perp$ does the job. In these two cases, there are infinitely many such solutions. Finally if $\dim V = 3$, then $V_1 = \operatorname{span}\{e_i\}$, $V_2 = \operatorname{span}\{e_i, Be_i\}$ and $V_3 = V$ is a flag which implements the tridiagonalization.

Indeed condition (ii) guarantees that the vector u in the solution of (1) and (2) has four non-zero coordinates. The proof shows that the tridiagonalizing flag consists of $V_1 = \mathbb{C}u$, $V_2 = \operatorname{span}\{u, Au\}$ and $V_3 = \operatorname{span}\{u, Au, A^2u\}$. It follows that at least three coordinates are non-zero, for otherwise V_3 would not be three dimensional. However if say $u_4 = 0$, then $V_3 = \operatorname{span}\{e_1, e_2, e_3\}$. Thus the fourth basis vector in the tridiagonalization is e_4. Reversing the flag yields a tridiagonalization beginning with e_4. Thus the subspace $\operatorname{span}\{e_4, Be_4, B^2e_4, ABe_4\}$ is three dimensional, which contradicts condition (ii).

3. Genericity

In this section, we provide additional arguments to demonstrate that the 12 solutions found for Pati's Theorem are generically multiplicity one. We suspect that there should be a rather quick and straightforward way to see this, but we have not found such a method. Indeed our arguments are somewhat involved. For this reason, we have separated it off as a separate section so as not to distract from our short and illuminating proof of the main result.

Theorem 3.1. *There is a proper (real) subvariety of the set of Hermitian pairs (A, B) which contains all pairs which do not have exactly twelve tridiagonalizing flags.*

Proof. We will show that a root of multiplicity two or more will have to satisfy an additional polynomial equation. We will then consider the whole system as equations for arbitrary complex matrices A and B in order to avail ourselves of the results of complex algebraic geometry. The main tool is elimination theory to obtain a proper variety of pairs of matrices satisfying the whole system.

Consider a non-extraneous solution of our equations (1) and (2), say $[s_0 : t_0 : \lambda_0]$ and $u_0 = [w_0 : x_0 : y_0 : z_0]$. Thus we have $t_0 \neq 0$; and we may suppose that $w_0 \neq 0$ (either by rearranging coordinates or by using the observation in Remark 2.3). We have

$$(s_0A + t_0B - \lambda_0I)u_0 = 0$$
$$\det(u_0, Au_0, A^2u_0, BAu_0) = 0$$

Let $\mathcal{R}$ be the polynomial ring localized over the point $[s_0 : t_0 : \lambda_0]$, $u_0 = [w_0 : x_0 : y_0 : z_0]$; and let $\mathcal{M}$ be the unique maximal ideal. For definiteness, we hold $t = t_0$ and $w = w_0$ fixed. Then

$$\mathcal{M} = \langle s - s_0, \lambda - \lambda_0, x - x_0, y - y_0, z - z_0\rangle.$$

Let $\mathcal{J}$ be the ideal generated by equations (1) and (2) in $\mathcal{R}$. By Nakayama's Lemma, it suffices to show that $\mathcal{J} + \mathcal{M}^2 = \mathcal{M}$ in order to conclude that $\mathcal{J} = \mathcal{M}$. So we may work modulo $\mathcal{M}^2$, which contains all quadratic terms.

Consider (1) first. Equivalence is modulo $\mathcal{M}^2$.

$$\begin{aligned}(sA+tB-\lambda I)u &= (sA+tB-\lambda I)u - (s_0A+t_0B-\lambda_0I)u_0 \\ &= (sA+tB-\lambda I)(u-u_0) + \big((s-s_0)A-(\lambda-\lambda_0)I\big)u_0 \\ &\equiv (s_0A+t_0B-\lambda_0I)(u-u_0) + \big((s-s_0)A-(\lambda-\lambda_0)I\big)u_0\end{aligned}$$

This yields four linear terms from the coefficients of

$$\begin{bmatrix}-u_0 & Au_0 & (s_0A+t_0B-\lambda_0I)P\end{bmatrix}\begin{bmatrix}\lambda-\lambda_0\\ s-s_0\\ x-x_0\\ y-y_0\\ z-z_0\end{bmatrix}$$

where P is the injection of $\mathbb{C}^3$ onto the last three coordinates of $\mathbb{C}^4$; i.e., $(s_0A+t_0B-\lambda_0I)P$ is the last three columns of the matrix $s_0A+t_0B-\lambda_0I$.

Now consider condition (2), again working modulo $\mathcal{M}^2$. We rewrite $\det(u, Au, A^2u, BAu)$ by replacing u by $(u-u_0)+u_0$ in all four locations.

Split the determinant into a sum of 16 terms and observe that eleven terms are of degree at least two in $u-u_0$, so lie in $\mathcal{M}^2$; and that the zeroth order term vanishes by (2). Thus

$$\begin{aligned}
&\det(u, Au, A^2u, BAu)\\
&\quad\equiv \det(u-u_0, Au_0, A^2u_0, BAu_0) + \det(u_0, A(u-u_0), A^2u_0, BAu_0)\\
&\quad\ + \det(u_0, Au_0, A^2(u-u_0), BAu_0) + \det(u_0, Au_0, A^2u_0, BA(u-u_0))\\
&\quad= \Delta_2(x-x_0) + \Delta_3(y-y_0) + \Delta_4(z-z_0)
\end{aligned}$$

where Δ_k is a homogeneous polynomial of degree 3 in u. (It is also degree 4 and degree 1 in the coefficients of A and B, respectively.)

This yields a fifth linear generator of $\mathcal{J}+\mathcal{M}^2$. Now $\mathcal{M}/\mathcal{M}^2$ is a five dimensional vector space. Hence in order for $\mathcal{J}+\mathcal{M}^2 = \mathcal{M}$, it is necessary and sufficient that these five linear terms be linear independent. That is, the failure of being multiplicity one is that we must satisfy the homogeneous determinant condition

$$\begin{vmatrix} -u_0 & Au_0 & (s_0A+t_0B-\lambda_0 I)P \\ 0 & 0 & \Delta_2 \quad \Delta_3 \quad \Delta_4 \end{vmatrix} = 0. \tag{3}$$

At this point, we wish to consider A and B as complex matrices even though we have been working with Hermitian matrices all along. Equations (1), (2) and (3) make perfect sense. We make the same generic restrictions:

(i) A has distinct eigenvalues $\alpha_1, \alpha_2, \alpha_3, \alpha_4$. Denote the corresponding eigenspacess by $\mathbb{C}v_i$.
(ii) $\{v_i, Bv_i, B^2v_i, ABv_i\}$ are linearly independent for $1 \le i \le 4$.

Under these circumstances, there are always solutions of (1) and (2) for $t = 0$, namely $u = v_i$ and $s = \alpha_i$. Bezout's Theorem applies and there are exactly 24 solutions when the genericity hypotheses hold. Lemma 2.2 goes through unchanged, showing that the extraneous solutions account for precisely 12 of the solutions; leaving another 12. Indeed that proof is basis independent.

The extraneous solutions are also solutions of equation (3). We avoid this by using the fact that they all occur for $t = 0$, while the other solutions require $t \ne 0$. Thus we fix $t = 1$ and also $w := u_1 = 1$ in equations (1), (2) and (3) to obtain an affine system (1′), (2′) and (3′). (The condition that $w_0 \ne 0$ is clearly generic as well.)

These equations determine a solution ideal $\mathcal{I}$ and an associated variety $V = V(\mathcal{I})$ in the affine space $\mathbb{C}^2 \times \mathbb{C}^3 \times \mathcal{M}_4(\mathbb{C}) \times \mathcal{M}_4(\mathbb{C})$ corresponding to the variables (s,λ), (x,y,z), $\{a_{ij} : 1 \le i,j \le 4\}$ and $\{b_{ij} : 1 \le i,j \le 4\}$.

Since ($1'$) is really four equations, we have six affine equations. There are five variables, $(s, \lambda), (x, y, z)$, that we wish to eliminate. This is accomplished by elimination theory. Specifically the Closure Theorem [1, Theorem 5.6.1]) applies. Consider the projection $\pi(V)$ of V into $\mathcal{M}_4(\mathbb{C}) \times \mathcal{M}_4(\mathbb{C})$. The Zariski closure of $\pi(V)$ is the variety $V_1 = V(\mathcal{I}_1)$ where $\mathcal{I}_1 = \mathcal{I} \cap \mathbb{C}[\{a_{ij}, b_{ij}\}]$. The Closure Theorem shows that there is a proper subvariety $W \subset V_1$ so that $V_1 - W \subset \pi(V) \subset V_1$.

It remains to show that V_1 is a proper variety. All pairs (A, B) for which there are not exactly 12 tridiagonaling flags are all contained in $\pi(V)$. In Example 5.1, we exhibit a pair (A, B) where we compute twelve distinct solutions. Thus each has multiplicity one. Moreover at these solutions, equation ($3'$) is non-zero. By continuity, this persists in an open neighbourhood of the point (A, B). So $\pi(V)$ omits an open set. If V_1 were the whole space, the complement of $\pi(V)$ would be contained in the proper subvariety W. Therefore V_1 is proper.

Finally, since our example consisted of two Hermitian matrices, we see that V_1 has proper intersection V_1^r with the real variety of pairs of Hermitian matrices. ∎

4. Consequences of Pati's Theorem

In this section, we show that the tridiagonalizing unitary matrices can be chosen to have a special form if we make some sensible initial choices for representing A and B.

Recall that the unitary group $\mathrm{U}(n)$ contains the subgroup $\mathrm{O}(n)$ of real orthogonal matrices. Also if we fix a basis $\{e_i : 1 \le i \le n\}$, let $\mathcal{D}_n$ denote the algebra of diagonal matrices with respect to this basis. The intersection of the unitary group with $\mathcal{D}_n$ is an n-torus, which we denote by T^n when there is no ambiguity about the basis being used. We write P' to denote the transpose of a matrix P.

Corollary 4.1. *Let A and B be Hermitian 4×4 matrices with A diagonal. Then there exists a unitary matrix $U = PD$, with $P \in \mathrm{O}(4)$ real orthogonal and D a diagonal unitary, such that UAU^* $(= PAP')$ and UBU^* are tridiagonal matrices.*

Proof. It suffices to prove the corollary in the generic case in which A has four distinct eigenvalues, which we assume to be the case. By Theorem 1.1, there exists a unitary matrix U such that $S := UAU^*$ and $T := UBU^*$ are tridiagonal. Moreover, by multiplying U on the left by a diagonal unitary, we may assume U is chosen so that S is real. Then the equation $UA = SU$ implies that $\overline{U}A = S\overline{U}$, where $\overline{U}$ is the entrywise complex conjugate of U. Thus A and $U^{-1}\overline{U}$ commute. As A

is diagonal with distinct diagonal entries, we have $U^{-1}\overline{U} = E$, where E is a unitary diagonal matrix. Factor $E = D^{-2}$ where D is also a unitary diagonal matrix. It is now easy to check that $P := UD^{-1} \in \mathrm{O}(4)$ and $U = PD$. ∎

The following immediate corollary is of independent interest. We will provide a second proof that does not rely on Pati's Theorem.

Corollary 4.2. *If A is a Hermitian 4×4 matrix, then there exists a unitary matrix $U = PD$, with P real orthogonal and D diagonal, such that UAU^* is a tridiagonal matrix.*

Proof. We choose a unitary diagonal matrix D such that all entries in the first row (and column) of DAD^* are real. Then $DAD^* = B + iC$, where B is real symmetric and C is real skew-symmetric with zero first row (and column). It suffices to show that there exists a real orthogonal matrix P such that PBP^{-1} and PCP^{-1} are simultaneously tridiagonal. We may assume that $C \neq 0$, and thus has rank 2. Clearly, we may normalize C and conjugate by a real orthogonal matrix so that

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

This is a rotation of the plane spanned by first two basis vectors through a right angle. With respect to this basis, write

$$B = \begin{bmatrix} X & Y \\ Y' & Z \end{bmatrix}$$

where X and Z are real symmetric.

Observe that the centralizer of C in $\mathrm{O}(4)$ is $\mathrm{R}(2) \oplus \mathrm{O}(2)$ where $\mathrm{R}(2)$ is the group of rotations in the plane. If Y has rank at most one, we may conjugate B by an element of this group so that Y has 0 coefficients except in the $(2, 1)$ corner. This tridiagonalizes B and C. So we may suppose that Y has rank 2.

Instead, after conjugating B by an element of the centralizer of C, we may assume that X and Z are diagonal. Thus without changing C, we arrive at

$$B = \begin{bmatrix} a_1 & 0 & b_1 & b_2 \\ 0 & a_2 & b_3 & b_4 \\ b_1 & b_3 & a_3 & 0 \\ b_2 & b_4 & 0 & a_4 \end{bmatrix}$$

From now on, we work strictly in $\mathcal{M}_4(\mathbb{R})$ acting on $\mathbb{R}^4$. We will find a flag of real subspaces $V_1 \subset V_2 \subset V_3 \subset \mathbb{R}^4$ so that $BV_i + CV_i \subset V_{i+1}$

for $i = 1, 2$. This will tridiagonalize B and C. Moreover choosing an orthonormal basis $e_i \in V_i \ominus V_{i-1}$ yields vectors in $\mathbb{R}^4$ and thus the unitary $P = [e_1\, e_2\, e_3\, e_4]$ will be a real orthogonal matrix which implements the similarity.

We look for a unit vector $u = (0, 0, x, y)'$. Clearly

$$V_2 = \operatorname{span}\{u, Bu, Cu\} = \operatorname{span}\{u, Bu\},$$

which is 2-dimensional because of the injectivity of Y. To tridiagonalize B and C, it suffices that

$$V_3 = \operatorname{span}\{u, Bu, CBu, B^2u\}$$

be 3 dimensional. This is equivalent to finding x and y so that

$$\det(u, Bu, CBu, B^2u) = 0.$$

A computation shows that this becomes

$$\begin{vmatrix} 0 & b_1x + b_2y & b_3x + b_4y & (a_1 + a_3)b_1x + (a_1 + a_4)b_2y \\ 0 & b_3x + b_4y & -b_1x - b_2y & (a_2 + a_3)b_3x + (a_2 + a_4)b_4y \\ x & a_3x & 0 & (b_1^2 + b_3^2 + a_3^2)x + (b_1b_2 + b_3b_4)y \\ y & a_4y & 0 & (b_1b_2 + b_3b_4)x + (b_2^2 + b_4^2 + a_4^2)y \end{vmatrix} = 0.$$

This is a homogeneous polynomial in x and y of degree 4. The coefficients of x^4 and y^4 are

$$-(b_1^2 + b_3^2)(b_1b_2 + b_3b_4) \quad \text{and} \quad (b_2^2 + b_4^2)(b_1b_2 + b_3b_4).$$

If $b_1b_2 + b_3b_4 \neq 0$, these terms have opposite signs. Therefore it follows from the Intermediate Value Theorem that this equation has a real solution $x = \cos\theta$ and $y = \sin\theta$ for some $\theta \in [0, \frac{\pi}{2}]$. If $b_1b_2 + b_3b_4 = 0$, then it is easily seen that $x = 1$ and $y = 0$ is a solution. This yields a joint tridiagonalization of B and C. ∎

5. The Real Case

As mentioned in the introduction, the second author and MacDonald [2] provide examples of real matrices which cannot be tridiagonalized using real orthogonal matrices. In the real case, the imaginary part of $T \in \mathcal{M}_4(\mathbb{R})$ has imaginary entries, and instead one is led to consider $T = A+B$ where $A = (T+T')/2$ and $B = (T-T')/2$. So B is skew Hermitian. However the formulation in terms of tridiagonalizing two Hermitian matrices also has a real version, namely a pair A, B of real symmetric matrices.

Unfortunately the answer in this case is also negative. We were able to ascertain by computer calculation that there are pairs of real symmetric matrices which cannot be tridiagonalized by real orthogonal matrices. The method was to use Maple to calculate all 12 solutions to the system of equations (1) and (2) and observe that they are not real. More precisely, if U^*AU and U^*BU are tridiagonal, the unitary U can be multiplied on the right by any diagonal unitary. There will be no real choice if the entries in some column do not all have the same argument.

When A and B are real symmetric and U is a unitary matrix such that U^*AU and U^*BU are tridiagonal, then $\overline{U}^* A\overline{U}$ and $\overline{U}^* B\overline{U}$ are also tridiagonal when $\overline{U}$ is the entrywise complex conjugate of U. Also one can flip these matrices by multiplying U on the left by the permutation that interchanges the first and fourth basis vectors and the second and third. So typically, the solutions will come in three sets of four.

Example 5.1. The first counterexample found by random testing was:

$$A = \begin{bmatrix} 3 & 0 & 0 & 0 \\ 0 & -6 & 0 & 0 \\ 0 & 0 & -5 & 0 \\ 0 & 0 & 0 & -9 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 6 & 1 & 4 & 1 \\ 1 & -1 & 3 & 3 \\ 4 & 3 & -3 & -1 \\ 1 & 3 & -1 & -1 \end{bmatrix}$$

The unitaries come in the four sets mentioned above, so we consider only one from each set of four. We need to check that U cannot be multiplied on the right by a diagonal unitary to make it real. As the vector u arising from the solution of our homogeneous equations is the first column of U, and we have normalized so that the first coordinate is 1, it is sufficient to observe that there are also non-real entries. The three vectors which arise as solutions are:

$$\begin{bmatrix} 1.0000000000 \\ -0.3253922499 + 0.3023748224i \\ 0.1925852419 + 0.5931640341i \\ -0.0317025305 - 0.0316957446i \end{bmatrix}$$

$$\begin{bmatrix} 1.0000000000 \\ 0.0705398360 - 1.757627576i \\ 1.157011674 - 1.296852842i \\ -0.9704272563 + 0.4830667032i \end{bmatrix}$$

and

$$\begin{bmatrix} 1.0000000000 \\ -3.266236362 - 2.634309378i \\ 0.4090422888 - 2.287361925i \\ -3.291124256 + 1.515089643i \end{bmatrix}$$

Example 5.2. A related question that could have been resolved in [2] is whether any pair of skew Hermitian matrices can be tridiagonalized by a real orthogonal matrix. The answer is no. Consider

$$A = \begin{bmatrix} R & 0 \\ 0 & R \end{bmatrix} \qquad B = \begin{bmatrix} 0 & -I \\ I & 0 \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} R & 0 \\ 0 & -R \end{bmatrix}$$

where $R = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. A simple computation shows that the only tridiagonal matrices which are orthogonally similar to B are $\pm A$ and $\pm C$. However $\pm A$ and $\pm C$ commute, while A and B do not. So A and B cannot be simultaneously tridiagonalized.

6. The Case of Symplectic Group

Let G be a connected almost simple complex Lie group and $\mathfrak{g}$ its Lie algebra. Fix a Cartan decomposition $\mathfrak{g} = \mathfrak{g}_u \oplus i\mathfrak{g}_u$ of $\mathfrak{g}$ and let G_u be the maximal compact subgroup of G corresponding to $\mathfrak{g}_u$. Also fix a maximal torus $T_u \subset G_u$ and let $T \subset G$ be its complexification. Then the Lie algebra $\mathfrak{h}$ of T is a Cartan subalgebra of $\mathfrak{g}$ and $\mathfrak{h}_u := \mathfrak{h} \cap \mathfrak{g}_u$ is the Lie algebra of T_u.

Let Σ be the root system of $\mathfrak{g}$ with respect to $\mathfrak{h}$ and fix its base (a set of simple roots), say, $\Pi = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$, where r is the rank of G. By analogy with the case of matrices, we say that an element $X \in \mathfrak{g}$ is *tridiagonal* if it belongs to the subspace

$$\mathfrak{h} \oplus \sum_{\alpha \in \Pi \cup -\Pi} \mathfrak{g}^\alpha,$$

where $\mathfrak{g}^\alpha$ is the (one-dimensional) root space of $\mathfrak{g}$ attached to α.

The group G_u acts on $\mathfrak{g}$ via the restriction of the adjoint representation of G. We say that an element $X \in \mathfrak{g}$ is *tridiagonalizable* if $g \cdot X$ is tridiagonal for some $g \in G_u$. With this terminology, Pati's Theorem says that in the case $G = \mathrm{SL}_4(\mathbb{C})$, every $X \in \mathfrak{sl}_4(\mathbb{C})$ is tridiagonalizable by an element of $\mathrm{SU}(4)$. This new definition of tridiagonalizability agrees with the old one in this case.

It is now natural to ask for which groups G is it true that all elements $X \in \mathfrak{g}$ are tridiagonalizable. As in the unitary case, an obvious necessary condition arises from dimension considerations. The space $\mathfrak{g}$ has dimension $2 \dim \mathfrak{g}_u$. The variety $\mathcal{T}$ of tridiagonal elements is easily seen to have complex dimension $3r$ and thus real dimension $6r$. Observe that every tridiagonal $X \in \mathfrak{g}$ can be transformed by T_u to another tridiagonal element whose components in the root spaces $\mathfrak{g}^\alpha$ with $\alpha \in \Pi$ are real multiples of fixed root vectors. This real subvariety $\mathcal{T}_r$ has real dimension $5r$. Hence a necessary condition for universal tridiagonalization is that

$$\dim \mathfrak{g} \leq \dim \mathfrak{g}_u + \dim \mathcal{T}_r = \dim \mathfrak{g}_u + 5r.$$

Therefore we require that

$$\dim(\mathfrak{g}_u) \leq 5r.$$

This is a very restrictive inequality. It is only valid for groups of type $A_1, A_2, A_3(= D_3)$ and $B_2(= C_2)$.

Cases A_1 and A_2 correspond to tridiagonalizing 2×2 and 3×3 complex matrices, which is very easy. If G is of type A_3 or C_2, then the equality holds in the dimension calculation, which makes the situation delicate. Type A_3 is handled by Pati's Theorem.

The main objective of this section is to prove that in the case of the symplectic group $G = \mathrm{Sp}_4(\mathbb{C})$, a group of type C_2, the analog of Pati's Theorem is valid, i.e., every $X \in \mathfrak{g}$ is tridiagonalizable. In the spirit of this paper, we shall express this result also in matrix terms.

Let us define $G := \mathrm{Sp}_4(\mathbb{C})$ as the matrix group:

$$\mathrm{Sp}_4(\mathbb{C}) := \{g \in \mathrm{GL}_4(\mathbb{C}) : g'Jg = J\},$$

where

$$J := \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & -1 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{bmatrix}.$$

Then $G_u := \mathrm{Sp}_4(\mathbb{C}) \cap \mathrm{U}(4)$ is a maximal compact subgroup of G, usually denoted by $\mathrm{Sp}(2)$. Hence,

$$\mathrm{Sp}(2) := \{g \in \mathrm{U}(4) : g'Jg = J\}.$$

We take T_u to be the maximal torus of $\mathrm{Sp}(2)$ consisting of the diagonal matrices:

$$\mathrm{diag}(e^{i\theta}, e^{i\varphi}, e^{-i\varphi}, e^{-i\theta}).$$

The Cartan subalgebra $\mathfrak{h}$ of $\mathfrak{g} = \mathfrak{sp}_4(\mathbb{C})$ consists of diagonal matrices

$$\mathrm{diag}(z, w, -w, -z), \quad z, w \in \mathbb{C}.$$

The Lie algebra $\mathfrak{g} = \mathfrak{sp}_4(\mathbb{C})$ has the following simple matrix description: It consists of all matrices

$$X = \begin{bmatrix} a & b & x & y \\ c & d & z & x \\ u & v & -d & -b \\ w & u & -c & -a \end{bmatrix},$$

where $a, b, c, d, u, v, w, x, y, z$ are arbitrary complex numbers. Let Π be the base of the root system defined by the Borel subalgebra of upper triangular matrices in $\mathfrak{g}$. Then an element $X \in \mathfrak{g}$ is tridiagonal if and only if it is a tridiagonal matrix.

Observe that if $X \in \mathfrak{g}$ then $X^* \in \mathfrak{g}$ also. Hence, any $X \in \mathfrak{g}$ can be written as $X = A + iB$ with $A, B \in \mathfrak{g}$ both Hermitian.

Theorem 6.1. *If $X \in \mathfrak{sp}_4(\mathbb{C})$, then there exists $U \in \mathrm{Sp}(2)$ such that UXU^* is a tridiagonal matrix. Equivalently, if $A, B \in \mathfrak{sp}_4(\mathbb{C})$ are Hermitian matrices, then there exists $U \in \mathrm{Sp}(2)$ such that U^*AU and U^*BU are tridiagonal.*

Proof. By Pati's Theorem there exists $U \in \mathrm{U}(4)$ such that U^*AU and U^*BU are tridiagonal matrices. We have to show that U can be chosen in $\mathrm{Sp}(2)$. Clearly, it suffices to prove this in the generic case.

Let us write (x, y) for the complex skew-symmetric bilinear product defined by the matrix J, i.e.,

$$(x, y) = x'Jy.$$

If $X \in \mathfrak{g}$, then we have $X'J = -JX$, which shows that

$$(Xx, y) = -(x, Xy)$$

holds for all $x, y \in \mathbb{C}^4$. In particular, $(x, x) = 0$ for every vector x.

From the proof of Theorem 2.1, we know that the first unit vector v_1 in a tridiagonalizing basis $\{v_1, v_2, v_3, v_4\}$ can be chosen as a scalar multiple of a nonzero vector u satisfying the system

$$(sA + B - \lambda I)u = 0, \quad \det(u, Au, A^2u, BAu) = 0.$$

Our first claim is that $(u, Au) = 0$.

In the generic case, the vectors u, Au and A^2u will be linearly independent, and thus $BAu = \alpha u + \beta Au + \gamma A^2u$ for some scalars α, β, γ. Observe that since $(u, A^2u) = -(Au, Au) = 0$,

$$(Bu, Au) = (\lambda u - sAu, Au) = \lambda(u, Au),$$

and

$$\begin{aligned}(Bu, Au) &= -(u, BAu)\\ &= -(u, \alpha u + \beta Au + \gamma A^2u)\\ &= -\beta(u, Au).\end{aligned}$$

The claim will now follow provided that $\lambda + \beta \neq 0$.

Assume now that $\lambda + \beta = 0$. Then we have

$$\begin{aligned}B^2u &= B(\lambda u - sAu)\\ &= \lambda(\lambda u - sAu) - s(\alpha u - \lambda Au + \gamma A^2u)\\ &= (\lambda^2 - \alpha s)u - \gamma sA^2u.\end{aligned}$$

Thus the vectors u, A^2u and B^2u are linearly dependent. This can be excluded by imposing additional genericity restrictions. Hence our claim is proved.

The vectors $\{v_1, v_2, v_3\}$ are obtained from $\{u, Au, A^2u\}$ using the Gram–Schmidt process. Since

$$(u, u) = (u, Au) = (u, A^2u) = 0,$$

it follows that $(u, v) = 0$ for all $v \in \text{span}\{u, Au, A^2u\}$. Therefore $(v_1, v_2) = 0 = (v_1, v_3)$. As the tridiagonalizing basis can be reversed, we also obtain that $(v_3, v_4) = 0 = (v_2, v_4)$.

Let $U = [v_1\, v_2\, v_3\, v_4]$ be the unitary obtained in this way which tridiagonalizes T. Then $U'JU$ is also unitary. It has matrix entries $v_i'Jv_j$, and in particular the only non-zero entries lie on the backward diagonal $i + j = 5$. We conclude that the complex numbers (v_1, v_4) and (v_2, v_3) have unit modulus. Hence by adjusting the phases of v_3 and v_4, we may assume that $(v_1, v_4) = (v_2, v_3) = 1$ and so $(v_4, v_1) = (v_3, v_2) = -1$. Hence our tridiagonalizing basis is not only orthonormal, but is also a symplectic basis. This means that U belongs to Sp(2). ∎

Remark 6.2. As $B_2 = C_2$ and $A_3 = D_3$, the tridiagonalization theorems can be restated in terms of complex orthogonal groups $\mathrm{SO}_5(\mathbb{C})$ and $\mathrm{SO}_6(\mathbb{C})$ and their suitable maximal compact subgroups (isomorphic to SO(5) and SO(6), respectively). We leave this routine task to the interested readers.

7. An Explicit rank Three

In this section, we provide an explicit (and simple) rank three 5×5 matrix which cannot be tridiagonalized. We use computer support to check various tedious calculations. But unlike Example 5.1, this can be checked by hand.

Remark 7.1. First we explicitly describe all of the irreducible components of the variety $\mathcal{D}_3\mathcal{T}$. They were found using Singular's **primdec** library [4]. It has 16 irreducible components, all with real dimension 18. These components are permuted by the symmetry group generated by the reflections in the two diagonals of the matrix X. Under this action, there are four orbits of size 2 and two orbits of size 4. As representatives of these orbits we can take the following varieties:

$$\mathcal{D}_3\mathcal{T}_1 : \; x_{11} = x_{12} = x_{54} = x_{55} = 0;$$

$$\mathcal{D}_3\mathcal{T}_2 : \; x_{23} = x_{54} = x_{55} = 0, \quad \begin{vmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{vmatrix} = 0\,;$$

$$\mathcal{D}_3\mathcal{T}_3:\ x_{23}=x_{43}=0,\quad \begin{vmatrix} x_{11} & x_{12}\\ x_{21} & x_{22}\end{vmatrix}=0,\quad \begin{vmatrix} x_{44} & x_{45}\\ x_{54} & x_{55}\end{vmatrix}=0\,;$$

$$\mathcal{D}_3\mathcal{T}_4:\ x_{34}=x_{54}=x_{55}=0,\quad \begin{vmatrix} x_{11} & x_{12} & 0\\ x_{21} & x_{22} & x_{23}\\ 0 & x_{32} & x_{33}\end{vmatrix}=0\,;$$

$$\mathcal{D}_3\mathcal{T}_5:\ x_{34}=x_{43}=0,\quad \begin{vmatrix} x_{11} & x_{12} & 0\\ x_{21} & x_{22} & x_{23}\\ 0 & x_{32} & x_{33}\end{vmatrix}=0,\quad \begin{vmatrix} x_{44} & x_{45}\\ x_{54} & x_{55}\end{vmatrix}=0\,;$$

$$\mathcal{D}_3\mathcal{T}_6:\ x_{45}=x_{54}=x_{55}=0,\quad \begin{vmatrix} x_{11} & x_{12} & 0 & 0\\ x_{21} & x_{22} & x_{23} & 0\\ 0 & x_{32} & x_{33} & x_{34}\\ 0 & 0 & x_{43} & x_{44}\end{vmatrix}=0\,.$$

7.2. **The Example.** Consider

$$T=\begin{bmatrix}1&1&1&1&1\\0&2&0&1&0\\0&0&3&1&1\\0&0&0&0&0\\0&0&0&0&0\end{bmatrix}\quad\text{and}\quad T^*=\begin{bmatrix}1&0&0&0&0\\1&2&0&0&0\\1&0&3&0&0\\1&1&1&0&0\\1&0&1&0&0\end{bmatrix}.$$

Denote the standard basis by e_1, e_2, e_3, e_4, e_5. Observe that $\sigma(T) = \{0, 1, 2, 3\}$. Let us write $Tx_i = ix_i$ and $T^*y_i = iy_i$ for $i = 1, 2, 3$ and let x_4, x_5 be a basis for $\ker T$. Note that $\ker T^* = \operatorname{span}\{e_4, e_5\}$ and $x_1 = e_1$. Then

$$x_2=\begin{bmatrix}1\\1\\0\\0\\0\end{bmatrix}\quad x_3=\begin{bmatrix}1\\0\\2\\0\\0\end{bmatrix}\quad x_4=\begin{bmatrix}1\\-1\\0\\2\\-2\end{bmatrix}\quad x_5=\begin{bmatrix}2\\0\\1\\0\\-3\end{bmatrix}\quad y_1=\begin{bmatrix}2\\-2\\-1\\-1\\1\end{bmatrix}\quad y_2=\begin{bmatrix}0\\2\\0\\1\\0\end{bmatrix}\quad y_3=\begin{bmatrix}0\\0\\3\\1\\1\end{bmatrix}$$

We will show that this is not tridiagonalizable by showing that it cannot be put into any of the forms of Remark 7.1. The easiest to eliminate is $\mathcal{D}_3\mathcal{T}_6$ because this form satisfies $\dim(\operatorname{Ran} A \vee \operatorname{Ran} A^*) = 4$; while $\dim(\operatorname{Ran} T \vee \operatorname{Ran} T^*) = 5$. Equally easy is $\mathcal{D}_3\mathcal{T}_5$ in which $A = A_1 \oplus A_2$ splits as an orthogonal direct sum of a 3×3 and a 2×2 matrix. The two blocks in such a decomposition would be spanned by eigenvectors of T and also of T^*. In particular, one would contain e_1 which can be easily seen to be cyclic for $\{T, T^*\}$.

Next consider cases $\mathcal{D}_3\mathcal{T}_1$, $\mathcal{D}_3\mathcal{T}_2$ and $\mathcal{D}_3\mathcal{T}_4$. They all have the condition $x_{54} = x_{55} = 0$. Suppose that T is tridiagonalized by a basis v_1, v_2, v_3, v_4, v_5. Then in these forms, v_5 is in the kernel of T^*. Then v_4 is a multiple of Tv_5 because $\operatorname{Ran} T$ is orthogonal to $\ker T^*$. And we require that $\dim \operatorname{span}\{v_5, v_4, Tv_4, T^*v_4\} = 3$. Set $v_5 = se_4 + te_5$. The following matrix must be rank 3:

$$\begin{bmatrix} 0 & s+t & 3s+2t & s+t \\ 0 & s & 2s & 3s+t \\ 0 & s+t & 3s+3t & 4s+4t \\ s & 0 & 0 & 3s+2t \\ t & 0 & 0 & 2s+2t \end{bmatrix}$$

However this matrix always has rank four. One way to determine this is to compute the determinant after deleting the last or second last row, and then taking the gcd. This is 1, and thus the rank is 4, eliminating these cases.

We must similarly dispense with the dual version where $v_5 \in \ker T$. As above, consider $v_5 = sx_4 + tx_5$ and look for s and t so that the subspace $\operatorname{span}\{v_5, T^*v_5, T^{*2}v_5, TT^*v_5\}$ has dimension 3. This leads to

$$\begin{bmatrix} s+2t & s+2t & s+2t & 2s+15t \\ -s & -s+2t & -s+6t & -2s+7t \\ t & s+5t & 4s+17t & 4s+21t \\ 2s & 3t & s+9t & 0 \\ -2s-3t & s+3t & 2s+7t & 0 \end{bmatrix}$$

Again this always has rank 4. Computations were made both with Maple and Singular.

The remaining case is $\mathcal{D}_3\mathcal{T}_3$. Here v_3 is an eigenvector of T. Moreover the subspaces $\operatorname{span}\{v_1, v_2\}$ and $\operatorname{span}\{v_4, v_5\}$ are invariant for T^* and the restrictions are rank 1. This means that they are each spanned by a vector in $\ker T^*$ and one of the eigenvectors $\{y_1, y_2, y_3\}$ of T^*. In particular, this would require that two of these three eigenvectors are orthogonal, which is evidently not the case. The dual version would require two of $\{x_1, x_2, x_3\}$ to be orthogonal. And this is not true either. This eliminates this case.

Hence T cannot be tridiagonalized.

References

[1] D. Cox, J. Little and D. O'Shea, *Ideals, varieties and algorithms*, Springer Verlag, New York, 1997.

[2] D.Ž. Đoković and M.L. MacDonald, *Orthogonal invariants of a matrix of order four and applications*, J. Pure Appl. Alg., to appear.

[3] C.K. Fong and P.Y. Wu,, *Band diagonal operators*, Lin. Alg. Appl. **248** (1996), 195–204.
[4] G.-M. Greuel and G. Pfister, *A Singular Introduction to Commutative Algebra*, Springer-Verlag, Berlin, 2002.
[5] J. Harris, *Algebraic Geometry, A First Course*, Springer–Verlag, New York, 1992.
[6] W.E. Longstaff, *On tridiagonalization of matrices*, Lin. Alg. Appl. **109** (1988), 153–163.
[7] V. Pati, *Unitary tridiagonalization in* $M(4,\mathbb{C})$, Proc. Indian Acad. Sci. (Math. Sci.) **111** (2001), 381–397.
[8] I.R. Shafarevich, *Basic algebraic geometry*, vol.1, 2nd edition, Springer–Verlag, Berlin, 1994.

Pure Math. Dept., U. Waterloo, Waterloo, ON N2L–3G1, CANADA
E-mail address: `krdavids@uwaterloo.ca`

Pure Math. Dept., U. Waterloo, Waterloo, ON N2L–3G1, CANADA
E-mail address: `djokovic@uwaterloo.ca`