ON COMMUTATORS OF COMPACT OPERATORS VIA BLOCK TRIDIAGONALIZATION: GENERALIZATIONS AND LIMITATIONS OF ANDERSON’S APPROACH

JIREH LOREAUX, SASMITA PATNAIK*, SRDJAN PETROVIC, AND GARY WEISS**

Abstract. We offer a new perspective and some advances on the 1971 Pearcy–Topping problem: Is every compact operator a commutator of compact operators? Our goal is to analyze and generalize the 1970’s work in this area of Joel Anderson combined with the work of the last named author of this paper. We reduce the general problem to a simpler sequence of finite matrix equations with norm constraints, while at the same time developing strategies for counterexamples. Our approach is to ask which compact operators $T$ are commutators $AB - BA$ of compact operators $A, B$; and to analyze the implications of Joel Anderson’s contributions to this problem, which will yield a generalization of his method. By extending the techniques of Anderson [1] we obtain new classes of operators that are commutators of compact operators beyond those obtained in [17] and [2]. And by employing the techniques of the last named author [22], we found obstructions to extending Anderson’s techniques in terms of certain constraints for $T$, with special focus on when $T$ is a strictly positive compact diagonal operator. Some of these constraints involve general universal block tridiagonal matrix forms for operators, and some involve $\mathcal{B}(\mathcal{H})$-ideal constraints. And in terms of these matrix forms, we give some equivalences, some sufficient conditions and some necessary conditions for this Pearcy–Topping problem to hold true. These matrix forms are a sparsification of matrix representations of an operator (an increase in the proportion of zeros in its corners by a change of basis) and we measure the support density of these forms. And finally we provide some necessary conditions for the Pearcy–Topping problem involving singular numbers and $\mathcal{B}(\mathcal{H})$-ideal constraints.

1. Introduction and Historical Survey

The classical commutator problem asks which $\mathcal{B}(\mathcal{H})$ operators are commutators, i.e., operators of the form $[A, B] = AB - BA$, for $\mathcal{H}$ an infinite dimensional, separable, complex Hilbert space. This was settled by Brown and Pearcy [3]: An operator is a commutator if and only if it is not a thin operator, i.e., not of the form $\lambda + K$, where $\lambda$ is a nonzero complex number and $K$ is compact. Following the Brown–Pearcy characterization of $\mathcal{B}(\mathcal{H})$-commutators [3], much of the research on single commutators was focused on special situations when operators $A, B$ belong to various ideals. Later research mostly focused on commutator ideals in $\mathcal{B}(\mathcal{H})$, $[\mathcal{I}, \mathcal{J}]$, i.e., the finite linear span of commutators of operators from $\mathcal{I}$ with operators from $\mathcal{J}$, in part because of the connection of $[\mathcal{I}, \mathcal{B}(\mathcal{H})]$ to traces. (Chronologically, see [19, 21, 1, 22, 14, 6, 13, 7].) The general characterization of commutator ideals $[\mathcal{I}, \mathcal{J}]$, was obtained in [6, Theorem 5.6] (or more simply distilled from [6, Introduction pp. 2–3 and remark following]): a normal operator $T \in \mathcal{I} \mathcal{J}$ belongs to $[\mathcal{I}, \mathcal{J}]$ if and only if $T$ belongs to $[\mathcal{I} \mathcal{J}, \mathcal{B}(\mathcal{H})]$ if and only if
$$\left| \frac{\lambda_1 + \lambda_2 + \cdots + \lambda_n}{n} \right| = O(\mu_n).$$

Here $\mathcal{I}, \mathcal{J}$ are arbitrary $\mathcal{B}(\mathcal{H})$-ideals at least one of which is proper, $\nu_n, \mu_n$ are monotone sequences of nonnegative numbers converging to 0 where $\text{diag}(\nu_n), \text{diag}(\mu_n)$ belong to the product ideal $\mathcal{I} \mathcal{J}$, and $T$ has eigenvalue sequence $(\lambda_n)$, the nonzero ones counted according to their multiplicities but taken in any

*Supported by Science and Engineering Research Board, Core Research Grant 002514
**Partially supported by Simons Foundation collaboration grants 245014 and 636554.
order subject only to the condition that $|\lambda_n| = O(\nu_n)$. (The existence of such an order and sequence $\nu$ is equivalent to the condition $T \in IJ$, which follows easily from the methods in [6].) A useful version of this characterization can be stated for any positive operator $T \in IJ$ using its $s$-numbers (singular value sequence): a positive $T$ belongs to $[I, J]$ if and only if the diagonal operator

$$\text{diag}\left(\frac{s_1 + \cdots + s_n}{n}\right) \in IJ.$$  

The fact that $s$-numbers are positive while eigenvalues may not be was important in [6] and [13] and related work of the authors of [13], primarily because working on positive cones, meaning cones of positive operators was essential for some of their results.

The single commutator problem for compact operators is another commutator question posed by Pearcy and Topping in 1971 which at the time of this writing remains open.

**Question 1.** Is every compact operator a commutator of compact operators?

Seminal results on this and related questions appeared in the 1970’s by Anderson [1], results by the last named author of this paper [22 settling [19] Problem 3(i)], and a natural test question posed by him. Motivated by Anderson’s rank one projection result in [1], he asked: Can a strictly positive compact operator be a commutator of compact operators? There followed 30 years later an affirmative solution to this test case in the unpublished work of Davidson–Marcoux–Radjavi [5] and independently a little later, similar results of Beltita joint with the second and fourth authors herein [2, 17]. A little earlier, in 1994, L. Brown, using block Hessenberg forms, had proved in [4] that no finite rank nonzero trace operator is a commutator of operators in Schatten classes $C_p$ and $C_q$, when $1/p + 1/q \geq 1/2$. (In particular, there are operators in $C_2$ that are not commutators of $C_4$ operators). Then in [6] Section 7] the authors in 2004 obtained a sufficient condition in terms of ideals $I, J$ in $B(H)$ for an operator to be a single commutator of operators from $I$ and $J$ (we denote this class by $[I, J]_1$) in the commutator ideal $[I, J]$. And recently in [7], K. Dykema and A. Krishnaswamy-Usha proved that all nilpotent compact operators are commutators of compact operators. From this, focus began on trying to represent specific compact operators $T$ and classes of compact operators as special kinds of commutators. Constructions for these, it was hoped, would lead to a more general strategy which began with [5, 17, 2] chronologically.

This paper focuses mainly on representing normal compact operators (i.e., diagonal compact operators), and particularly strictly positive ones, as commutators of compact operators. The latter are notoriously hard to represent, and the only known diagonal sequence patterns obtainable have either infinitely many zeros (Anderson) or have each strictly positive eigenvalue repeated according to a certain rigid multiplicity structure (Davidson–Marcoux–Radjavi and Beltita–Patnaik–Weiss, see herein Theorems 2.6 and 2.7). One of the central accomplishments of this paper (Section 2) is a modification of these constructions leading to the affirmative answer to a question arising from [2, p. 9]: there are infinitely many strictly positive compact operators with distinct nonzero eigenvalues that are commutators of compact operators, thereby expanding the known class. We hope from these constructions more general ones might evolve, perhaps leading to a full characterization of which normal compact operators are commutators of compact operators.

Next we address the question whether Anderson’s commutator model [1] can be used to obtain every positive compact operator. In Section 3 we present several obstructions limiting this model. Some of them have the flavor of majorization type constraints involving arithmetic means arising in the work of Ky Fan [8].

To keep manageable the computations, following Anderson’s approach, it was natural to use sparse matrices (sparse meaning with most or at least many entries zero). For instance, the commutator of an upper and a lower weighted shift is a diagonal operator, and this phenomenon has led to many important consequences in the early literature. To sparsify a matrix (change basis to increase the zero regions) for a general $T$, and in particular for $T = AB - BA$, we construct an orthonormal basis for which the matrix
of $T$ is sparse and for which, when needed, simultaneously, the $A$ and $B$ matrix forms must also be sparse and of the same especially convenient form. And especially when $A, B$ are sparse of this convenient form, the computations of $AB$ and $BA$ become more manageable. A change of basis corresponds to a unitary operator $U$, which yields the matrix representation of $T$ in the new basis $\langle U e_n \rangle$ as $U^{-1} T U$ in the original basis $\langle e_n \rangle$ of $T$. This leads to the question: how sparse can $U^{-1} T U$ be made? In Section 4 we lift from [18] a brief description of certain matrix sparsifications. An in-depth study on this can be found there. Section 4 also provides a new Pearcy–Topping commutator problem equivalence (Theorem 4.6) based on this that in a sense fully generalizes Anderson’s approach. Historically, Anderson’s tour de force solved affirmatively the Pearcy–Topping test question that every rank one projection operator is a commutator of compact operators, and used this to prove that every compact operator is a commutator of a compact operator with a bounded operator. For this rank one projection case he employed block tridiagonal matrices with block sizes growing arithmetically. And subsequently motivated by Anderson’s model, our [22] introduced general staircase forms which led us to general block tridiagonal forms with further matrix sparsifications in our precursor [18] to this paper.

A word about evolving block tridiagonal terminology used herein. The displays (2.1)–(2.2) in Section 2 constitute Anderson’s model (AM). In Section 3 we generalize (AM) to an exponential Anderson’s model (EAM) by choosing instead exponentially growing block sizes whereas (AM) uses arithmetically growing block sizes. And after that we use our universal block tridiagonal forms which generalize both. The exponentially growing block sizes in our universal block tridiagonal forms apply to all operators and can be derived from any arbitrary basis.

In short, the organization of this paper is as follows. In Section 2 we reiterate the classical Anderson’s model that he used to obtain a rank one projection as a commutator of compact operators. Our main result (Theorem 2.10), expanding on Anderson’s model, is that with a controlled perturbation of his nonzero entries one obtains uncountably many diagonal compact operators with distinct strictly positive entries. Along with their unitary equivalents, this expands the known class and solves a problem arising from [2] that evolved from [1, 5, 17].

In Section 3 we show that not every compact diagonal operator can be obtained using Anderson’s construction. An obstruction comes from a necessary trace condition for Anderson’s model to work. We formulate this condition in two different ways, one of them having a flavor of majorization such as appears in Ky Fan’s theorem [8] (or see [11, Lemma 4.1]). We found that the arithmetic growth of block sizes of Anderson’s block tridiagonal form is a main culprit. Using our universal block tridiagonal forms which have exponential block growth sizes, we consider an expanded version of Anderson’s model using instead exponential growth size blocks to provide more degrees of freedom. But even when the blocks are chosen to grow exponentially, we prove that this particular construction offers no improvement.

The study of this exponential Anderson model ties in naturally with the content of Section 4 where we, in a sense, fully generalize Anderson’s model. We show that given an arbitrary finite set of operators and an arbitrary basis, one can derive from these a new basis so that all the operators simultaneously take block tridiagonal form with identical block sizes. And these central (and hence all) block sizes grow exponentially (Theorems 4.1, 4.3 and 4.4). As a consequence we obtain a block tridiagonal equivalence to the Pearcy–Topping compact commutator problem (Theorem 4.6). This result highlights the value of generalizing Anderson’s construction. But at the same time also explains in a sense its limitations. And explains the fact, established in Section 3 that the obstruction detected there remains in force when the finite diagonal matrix blocks (with arithmetic growth) of the block tridiagonal operators employed are instead full finite matrix blocks. (Observe that Anderson’s block tridiagonal matrix solutions possess merely two nonzero block diagonals. Whereas here we consider fully general block tridiagonal forms.) We present several modifications of the construction in Theorem 4.1 which, in addition to obtaining the block tridiagonal forms (Theorems 4.1, 4.3 and 4.4), guarantee that additionally some of the entries in
the off-diagonal blocks can further be made 0, i.e., one can achieve further sparsification (Theorems 4.9 and 4.14).

In Section 5 we make a rough comparison between the sparsity of the matrix forms discussed herein (their possible nonzero entries called the support entries) and that of other well-known canonical matrix forms for operators. We define the basis dependent concept of support density (a measure of sparsification), and we calculate it for various forms. In particular, we show that for our block tridiagonal forms used here and for the staircase forms developed by the last named author in [22], the support density is $2/3$, but that one of our modifications here of our original block tridiagonal forms reduce the support density to $1/2$ (see Theorems 5.1 and 5.3). In contrast, in the bases diagonalizing compact normal operators or having Anderson model forms, the support density is 0. And for triangular or Hessenberg matrices, the support density is $1/2$. But of course not all operators have bases that admit these forms (which for Hessenberg matrices and selfadjoint operators, can depend on the existence of a cyclic vector (see paragraphs preceding Theorem 5.1)). But all operators do have universal block tridiagonal forms in some basis. In fact, from any basis one can derive a new basis to achieve this.

Section 6 provides many necessary conditions for the Pearcy–Topping problem involving singular numbers (alias $s$-numbers) and algebraic $\mathcal{B}(\mathcal{H})$-ideal constraints.

2. Anderson’s Model (AM)

This section presents our first steps in extending the contributions of [1] and [22] beyond the previous preliminary work (listed chronologically) of [5,17,18]. We start with some beginning history, then proceed to solutions of some test problems that later arose from that history, namely, about which compact positive diagonal operators are commutators of compact operators.

Our main motivation for studying sparse matrices starting in [15] with a description in Section 4 herein, comes from the contributions of [1] and [22] to the 1971 questions of Pearcy and Topping [19]. In particular: (i) Is every compact operator a single commutator of compact operators; and (ii) Is every trace class, trace zero operator a commutator of Hilbert–Schmidt operators, or at least a finite sum of such commutators? These turned out to be the most important among several results and seminal open questions about commutators and their linear spans listed in [19]. For starters, they proposed a surprisingly simple test question which turned out to be groundbreaking: Is every rank one projection a single commutator of compact operators? An affirmative answer was provided by Anderson in [1] along with applications (Theorems 2.1 to 2.3 herein). Via weighted shifts, it is easy to obtain selfadjoint operators with initial sums of its eigenvalues converging to 0 as commutators of compact operators (see [9, Theorem 1]). It turned out, for nonzero positive compact operators (hence initial sums of its eigenvalues do not converge to 0), representing them as a commutator of compact operators was very hard. This is one of the reasons why targeting positive compact operators became so important.

**Theorem 2.1** (Anderson [1], 1977). For $P$ a rank one selfadjoint projection, there exist compact operators $C$ and $Z$ for which $P = CZ - ZC$.

An application of this to a more general commutator result is:

**Theorem 2.2** (Anderson [1], 1977). If $T$ is a compact operator with $\ker T$ containing an infinite dimensional reducing subspace of $T$, then there are compact operators $C$ and $Z$ for which $T = CZ - ZC$.

Regarding the Pearcy–Topping commutator problem for general compact operators, a weaker but completely general result is:

**Theorem 2.3** (Anderson [1], 1977). If $T$ is a compact operator, then there is a compact operator $C$ and a bounded operator $Z$ for which $T = CZ - ZC.$
Anderson’s Model (AM). Historically, many results about commutators employed matrices that were weighted shifts or their adjoints, possibly with block weights (replacing scalar weights). Anderson used the following block matrix forms, with central block sizes growing arithmetically:

\[
C = \begin{pmatrix}
0 & A_1 & 0 & \ldots \\
B_1 & 0 & A_2 & \ddots \\
0 & B_2 & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots
\end{pmatrix}, \quad Z = \begin{pmatrix}
0 & X_1 & 0 & \ldots \\
Y_1 & 0 & X_2 & \ddots \\
0 & Y_2 & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots
\end{pmatrix}.
\] (2.1)

The sizes of these diagonal square blocks grow arithmetically: \(1 \times 1, 2 \times 2, 3 \times 3, \ldots\). And the nonzero rectangular blocks grow arithmetically and are all shift-like, i.e., with nonzero entries only on the certain diagonals:

\[
A_n = \begin{pmatrix}
a_{1,n} & 0 & \cdots & 0 & 0 \\
0 & a_{2,n} & \ddots & \vdots & \vdots \\
\vdots & \ddots & \ddots & 0 & \vdots \\
0 & \cdots & 0 & a_{n,n} & 0
\end{pmatrix}, \quad X_n = \begin{pmatrix}
0 & x_{1,n} & 0 & \cdots & 0 \\
\vdots & 0 & x_{2,n} & \ddots & \vdots \\
\vdots & \vdots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & 0 & x_{n,n}
\end{pmatrix},
\]

\[
B_n = \begin{pmatrix}
0 & \cdots & \cdots & 0 \\
-b_{1,n} & 0 & \cdots & 0 \\
0 & -b_{2,n} & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & \cdots & 0 & -b_{n,n}
\end{pmatrix}, \quad Y_n = \begin{pmatrix}
y_{1,n} & 0 & \cdots & 0 \\
0 & y_{2,n} & \ddots & \vdots \\
\vdots & \ddots & \ddots & 0 \\
0 & 0 & \cdots & y_{n,n}
\end{pmatrix}.
\] (2.2)

Throughout the paper we will refer to operators \(C\) and \(Z\) described by (2.1)–(2.2) as the (classical) Anderson model (AM).

The specific matrix structure of the operators \(C\) and \(Z\) in (2.1) allowed Anderson to reduce the operator equation \(CZ - ZC = D\), for \(D\) block diagonal (in Anderson’s case \(D = P\) the rank one projection) to a system of block matrix equations. These equations can be split into two groups: those that ensure that \(CZ - ZC\) is block diagonal, and those that force those diagonal blocks to match the diagonal blocks \(\langle D_n \rangle\) of \(D\). These sets of equations are

\[
A_nX_{n+1} = X_nA_{n+1} \quad \text{and} \quad B_{n+1}Y_n = Y_{n+1}B_n,
\]

\[
D_1 = A_1Y_1 - X_1B_1, \quad D_{n+1} = B_nX_n - Y_nA_n + A_{n+1}Y_{n+1} - X_{n+1}B_{n+1}, \quad n = 1, 2, \ldots
\] (2.3)

Remark 2.4. To solve for \(D = CZ - ZC\) where \(D\) is block diagonal with central block sizes \(1 \times 1, 2 \times 2, 3 \times 3, \ldots\) and with \(C, Z\) having matrix forms (2.1), it is necessary and sufficient to simply solve the system (2.3) for these finite blocks, so that the \(C, Z\) finite block norms all tend to 0. (The latter condition ensures \(C, Z\) are compact, but moreover it is also a necessary condition by Theorem 6.8.)

Remark 2.5. Displays (2.1)–(2.2) being what we mean by Anderson’s model, in Section 3 we generalize (AM) to an exponential Anderson’s model (EAM) by choosing instead exponentially growing block sizes. And later we employ our universal block tridiagonal forms in Section 4 which generalize both by showing every operator in some basis is block tridiagonal with universal block sizes growing exponentially.

Based on Anderson’s findings above, the last named author asked in 1976 whether any strictly positive compact operator could be a commutator of compact operators (see [2, p. 2, Acknowledgement and prior paragraph]). This, based on staircase matrix form calculations from [22, Theorem 2 and lead up], in the hope of obtaining a class of counterexamples. And also because if the kernel of \(D\) has infinite
dimension, then $D$ is a commutator of compact operators by Anderson’s Theorem 2.2 above. But the finite dimensional kernel case for nonnegative $D$ remains unknown, in particular the zero dimensional kernel case, that is, the strictly positive case stated above. Of course, a positive solution to the strictly positive case trivially yields a positive solution to the general positive case.

It is natural to ask whether the construction that Anderson has employed can be modified to answer this commutator question or, more generally, for all compact normal operators $D$ (i.e., diagonal compact operators). One such advance came in [2].

**Theorem 2.6** ([2, Theorem 3.1]). Positive compact operators are commutators of compact operators when they have eigenvalue sequences:

$$\langle d_1, \frac{d_2 - d_1}{2}, \frac{d_3 - d_2}{3}, \frac{d_4 - d_3}{3}, \frac{d_5 - d_4}{3}, \ldots \rangle,$$

where $\langle d_n \rangle$ is a nondecreasing sequence of nonnegative numbers but $d_n/n \to 0$.

This was earlier proved by Davidson, Marcoux, and Radjavi [5], but the result was never published nor disseminated. But at the time in a private communication to the last named author of this paper they asked if knowing this result would lead to achieving all strictly positive diagonal operators being so representable as commutators of compact operators. We see below in Section 3 (Theorem 3.11 and Corollary 3.12) that not all strictly positive compact operators can be representable this way using Anderson’s model, nor by using an expanded version, the exponential Anderson’s model (EAM).

An equivalent formulation of Theorem 2.6 appears in [17].

**Theorem 2.7** ([17, Theorem 3.2.1]). Positive compact operators are commutators of compact operators when they have eigenvalue sequences:

$$\langle d_1, \frac{d_2}{2}, \frac{d_3}{3}, \frac{d_4}{3}, \frac{d_5}{3}, \frac{d_6}{3}, \ldots \rangle,$$

where $\frac{1}{n} \sum_{j=1}^{n} d_j \to 0$.

The proof of Theorem 2.6 is constructive, with a clever alteration of the Anderson’s matrices (2.2) by inserting a factor $\sqrt{d_n}$ in front of $A_n$, $X_n$, $B_n$ and $Y_n$.

Theorem 2.2 had a severe restriction that the target operator must have an infinite dimensional reducing subspace on which it is 0, while Theorem 2.6 (and Theorem 2.7) requires that the eigenvalues repeat. Therefore, it is of interest to ask whether for all positive or all normal compact operators one can move beyond these two patterns of diagonal sequences. In particular, is it possible to obtain a diagonal operator with the diagonal entries distinct positive numbers? Our main result in this section (Theorem 2.10) is that a modification of Anderson’s model yields an affirmative answer to this question.

The proof of this theorem relies on the following two lemmas. Throughout this paper we use the term positive number to mean strictly positive.

**Lemma 2.8.** Let $0 < L < 1$. Then there exists a sequence $\langle \alpha_n \rangle$ of positive numbers such that

$$\alpha_{n+1} \geq \left(1 + \frac{L}{n}\right) \alpha_n, \quad \text{and} \quad \frac{\alpha_n}{n} \to 0. \quad (2.4)$$

**Proof.** Let $L < M < 1$. We define the sequence $\langle \alpha_n \rangle$ inductively. Let $\alpha_1 > 0$. If $\alpha_1, \ldots, \alpha_n$ have been selected, we choose $\alpha_{n+1}$ so that

$$\left(1 + \frac{L}{n}\right) \alpha_n \leq \alpha_{n+1} \leq \left(1 + \frac{M}{n}\right) \alpha_n.$$
Then \( \langle \alpha_n \rangle \) satisfies the first condition in (2.4) and
\[
\alpha_{n+1} \leq \prod_{k=1}^{n} \left( 1 + \frac{M}{k} \right) \alpha_1.
\]

But it is well known that the product above is asymptotically equivalent to \( n^M \), (i.e., \( \prod_{k=1}^{n} \left( 1 + \frac{M}{k} \right)/n^M \) is bounded above and below via taking logarithms) so the second condition in (2.4) holds, and the lemma is proved.

**Lemma 2.9.** Let \( \langle \varepsilon_n \rangle_{n=1}^{\infty} \) be any bounded sequence of strictly positive numbers with \( L := \sup \varepsilon_n < 1 \), let \( \varepsilon_0 = 0 \), and let \( \langle \alpha_n \rangle_{n=1}^{\infty} \) be any sequence of positive numbers satisfying (2.4). Then there exists a sequence of strictly positive numbers \( \langle d_n \rangle_{n=1}^{\infty} \) with the following properties:

(i) \[ 0 < d_1 \leq \alpha_1. \] (2.5)

(ii) For all \( n \geq 1 \) and \( 1 \leq k \leq n+1 \)
\[ d_n \frac{1 + \varepsilon_{k-1}/n}{1 + \varepsilon_k/(n+2)} < d_{n+1} \leq \alpha_{n+1}. \] (2.6)

**Proof.** We will construct the sequence \( \langle d_n \rangle \) inductively. Let \( d_1 \) be any real number satisfying condition (2.5). Suppose that the numbers \( d_1, d_2, \ldots, d_n \) have been selected so that they satisfy conditions of the form (2.5)-(2.6). Notice that, in view of \( 0 < \varepsilon_n \leq L \) as well as the induction hypothesis, we have for \( n \geq 1 \) that
\[ d_n \frac{1 + \varepsilon_{k-1}/n}{1 + \varepsilon_k/(n+2)} < d_n \left( 1 + \frac{L}{n} \right) \leq \alpha_n \left( 1 + \frac{L}{n} \right) \leq \alpha_{n+1}. \]

Thus, \( d_{n+1} \) can be chosen anywhere in the interval
\[ \left( d_n \frac{1 + \varepsilon_{k-1}/n}{1 + \varepsilon_k/(n+2)}, \alpha_{n+1} \right). \] (2.7)

Now we can prove the promised theorem.

**Theorem 2.10.** Let \( \langle \varepsilon_n \rangle_{n=0}^{\infty} \), their supremum \( L \), and \( \langle \alpha_n \rangle_{n=1}^{\infty} \) be as in Lemma 2.9 and let \( \langle d_n \rangle_{n=1}^{\infty} \) be a sequence provided by this lemma.

Set \( d_{1,1} = d_1 + \varepsilon_1/2 \) and for all \( n \geq 1 \) and \( 1 \leq k \leq n+1 \) and set
\[
d_{k,n+1} = \frac{d_{n+1} - d_n}{n+1} + \frac{\varepsilon_k d_{n+1}}{(n+1)(n+2)} - \frac{\varepsilon_{k-1} d_n}{n(n+1)}. \] (2.8)

Then the diagonal operator \( D = \bigoplus D_n \), where \( D_n = \text{diag}(d_{k,n})_{k=1}^{n} \), is strictly positive and is a commutator of compact operators. Moreover, for many such \( D \), all the diagonal entries \( \langle d_{k,n} \rangle \) are distinct.

**Proof.** We first show that \( d_{k,n} > 0 \) for all \( 1 \leq k \leq n \). Indeed, \( d_{1,1} = d_1 + \varepsilon_1/2 > 0 \). And for \( n \geq 1 \) and \( 1 \leq k \leq n+1 \),
\[
d_{k,n+1} = \frac{d_{n+1} - d_n}{n+1} + \frac{\varepsilon_k d_{n+1}}{(n+1)(n+2)} - \frac{\varepsilon_{k-1} d_n}{n(n+1)} > 0
\]
is equivalent to the inequality
\[ d_{n+1} \left( 1 + \frac{\varepsilon_k}{n+2} \right) > d_n \left( 1 + \frac{\varepsilon_{k-1}}{n} \right) \]
which follows from the left-hand inequality in (2.6). Thus \( D \) is strictly positive.
In order to establish that $D$ is a commutator of compact operators, we will rely on the construction used in the proof of Theorem 2.7 but with a controlled perturbation of matrices in (2.2). Indeed, Anderson’s original parameters were (after setting his parameter $t = 1/2$)

$$a_{k,n} = \frac{\sqrt{n + 1 - k}}{n}, \quad x_{k,n} = \frac{\sqrt{k}}{n},$$

$$b_{k,n} = \frac{\sqrt{k}}{n + 1}, \quad y_{k,n} = \frac{\sqrt{n + 1 - k}}{n + 1}.$$  

We define the numbers:

$$a_{k,n} = \sqrt{d_n} \frac{\sqrt{n + 1 - k}}{n}, \quad x_{k,n} = \sqrt{d_n} \frac{\sqrt{k + \varepsilon_k}}{n},$$

$$b_{k,n} = \sqrt{d_n} \frac{\sqrt{k + \varepsilon_k}}{n + 1}, \quad y_{k,n} = \sqrt{d_n} \frac{\sqrt{n + 1 - k}}{n + 1},$$  \hspace{1cm} (2.9)

for $n \in \mathbb{N}$ and $1 \leq k \leq n$; the matrices $A_n, X_n, B_n, Y_n$ as in (2.2); the operators $C, Z$ as in (2.1). Next we show $C, Z$ are compact and $D = CZ - ZC$ for $D$ as defined above.

It is easy to verify that $C$ and $Z$ are compact. Indeed, the inequalities $|a_{k,n}|, |b_{k,n}|, |x_{k,n}|, |y_{k,n}| \leq \sqrt{d_n}/n$ are obvious and they imply that $\|A_n\|, \|B_n\|, \|X_n\|, \|Y_n\| \leq \sqrt{d_n}/n$. In view of the right-hand inequality in (2.6) and the second condition in (2.4), i.e., $0 \leq d_{n+1} \leq \alpha_{n+1}$ and $\alpha_n/n \to 0$, we obtain that $\|A_n\|, \|B_n\|, \|X_n\|, \|Y_n\| \to 0$ as $n \to \infty$, so $C$ and $Z$ are compact.

A calculation shows that off-diagonal blocks of $CZ - ZC$ are of the form $A_nX_n+1 - X_nA_n+1$ and $B_nB_n+1Y_n - Y_nB_nB_n+1$. It follows that they all vanish if and only if

$$a_{k,n}x_{k,n+1} = x_{k,n}a_{k+1,n+1} \quad \text{and} \quad b_{k,n+1}y_{k,n} = y_{k,n+1}b_{k,n+1},$$  \hspace{1cm} (2.10)

for $n \in \mathbb{N}$ and $1 \leq k \leq n$. A straightforward verification that all of these hold is left to the reader.

Since the off-diagonal blocks of $CZ - ZC$ are zero, $CZ - ZC$ is a block diagonal matrix. And using the fact that $C$ and $Z$ have the forms (2.1) it is straightforward to verify that the diagonal blocks of $D$ are given by the formulas (2.3)

$$D_{n+1} = A_{n+1}Y_{n+1} - X_{n+1}B_{n+1} + B_nX_n - Y_nA_n, \quad n \geq 0,$$  \hspace{1cm} (2.11)

with the convention that $A_0 = B_0 = X_0 = Y_0 = 0$. Thus, the first $(1 \times 1)$ block is

$$D_1 = A_1Y_1 - X_1B_1 = (a_{1,1}y_{1,1} + x_{1,1}b_{1,1}) = \left(\frac{d_1}{2} + \frac{d_1(1 + \varepsilon_1)}{2}\right) = (d_1 + \varepsilon_1/2) = (d_1,1).$$  \hspace{1cm} (2.12)

And using (2.11), it is an exercise in matrix algebra to show that, for $n \geq 1$, the $(n+1) \times (n+1)$ diagonal block of $CZ - ZC$ equals

$$D_{n+1} = \frac{d_{n+1} - d_n}{n + 1}I_{n+1} + \begin{pmatrix} \varepsilon_1d_{n+1}/(n+1)(n+2) & \varepsilon_2d_{n+1}/(n+1)(n+2) & \cdots & \varepsilon_{n+1}d_{n+1}/(n+1)(n+2) \\ \varepsilon_1d_{n+1}/(n+1)(n+2) & \varepsilon_2d_{n+1}/(n+1)(n+2) & \cdots & \varepsilon_{n+1}d_{n+1}/(n+1)(n+2) \\ \vdots & \vdots & \ddots & \vdots \\ \varepsilon_1d_{n+1}/(n+1)(n+2) & \varepsilon_2d_{n+1}/(n+1)(n+2) & \cdots & \varepsilon_{n+1}d_{n+1}/(n+1)(n+2) \end{pmatrix},$$

where the symbol $I_{n+1}$ denotes the $(n+1) \times (n+1)$ identity matrix. Thus each $(n+1) \times (n+1)$ diagonal block is itself a diagonal matrix. A comparison with (2.8) shows that these diagonal entries are exactly $(d_{k,n+1})_{k=1}^{n+1}$ which proves that $D$ is a commutator of compact operators.

Finally, we note that it is often the case that $d_{k,n}$ are all distinct. Indeed, notice that in the inductive construction in the proof of Lemma 2.9 it suffices to choose $d_{n+1}$ in the interval (2.7) so that none of the values $d_{k,n+1}$ (1 \leq k \leq n+1) coincide with the values $d_{j,m}$ (1 \leq j \leq m \leq n+1).$ Whenever $(\varepsilon_k)_{k=0}^{\infty}$ are themselves distinct, this is trivially possible. Indeed, since each $d_{k,n+1}$ is a differentiable function of $d_{n+1}$ for each $k$, with derivative $1/(n + 1 + \varepsilon_k)/((n+1)(n+2))$ and the $\varepsilon_k$ are distinct positive numbers,
there are at most finitely many choices \(((n + 1)(n + 2)/2)\) of \(d_{n+1}\) for which \(d_{k,n+1} = d_{j,n+1}\). Likewise, since each \(d_{k,n+1}\) is a strictly increasing function of \(d_{n+1}\), there are only finitely many choices of \(d_{n+1}\) for which \(d_{k,n+1} = d_{j,m}\) for some \(1 \leq j \leq m \leq n\) and \(1 \leq k \leq n+1\). Since there are only finitely many such values but infinitely many choices for \(d_{n+1}\) in the interval \([2.7]\), it is possible to choose all the \(d_{k,n}\) distinct. In fact, choosing each \(d_{n+1}\) in the interval \([2.7]\) uniformly at random (when the \(\varepsilon_k\) are all equal to 0) would result in distinct \(d_{k,n}\) with probability 1.

These techniques somewhat resemble partial results on the same distinctness problem appearing in the Davidson, Marcoux, and Radjavi preprint (private communication) \cite{5} mentioned earlier.

3. Anderson’s model and its generalization: obstructions and limitations

In view of the progress accomplished in Section\(\amalg\) it is natural to ask whether a clever choice of entries in the off-diagonal blocks of Anderson’s model could always yield an arbitrary positive diagonal operator as a commutator of compact operators. It turns out that this is not the case. Indeed, the arithmetic growth of his block sizes presents an obstruction (see Corollary \(\amalg\) and Theorem \(\amalg\)).

In this section, we will concern ourselves with the limitations of Anderson’s model by establishing necessary conditions on certain generalizations of his model. Throughout this section we will express \(C, Z\) in block matrix form with finite block sizes, that is,

\[
C = \begin{pmatrix}
C_1 \quad A_1 \quad * \quad \cdots \\
B_1 \quad C_2 \quad A_2 \quad \cdots \\
* \quad B_2 \quad C_3 \quad \cdots \\
\vdots \quad \cdots \quad \cdots \quad \cdots
\end{pmatrix}, \quad Z = \begin{pmatrix}
Z_1 \quad X_1 \quad * \quad \cdots \\
Y_1 \quad Z_2 \quad X_2 \quad \cdots \\
* \quad Y_2 \quad Z_3 \quad \cdots \\
\vdots \quad \cdots \quad \cdots \quad \cdots
\end{pmatrix}, \quad (3.1)
\]

with the central blocks \(C_n, Z_n\) of size \(k_n \times k_n\). We will be especially interested in the case when \(C\) (or alternatively \(Z\)) is in block tridiagonal form; that is, when the blocks of \(C\) (or alternatively \(Z\)) specified by asterisks in \(3.1\) are zero. This is pertinent for two reasons: (i) in Anderson’s model, both operators were in block tridiagonal form; (ii) in Section \(\amalg\) we show how to place any operator in block tridiagonal form with respect to some basis.

Our constraints will often be phrased in terms of norm conditions on the rectangular blocks \(A_n\) and \(B_n\). Of course, \(\|A_n\|\) is the usual operator norm, but we will also use the notation \(\|A_n\|_1\) to refer to the trace norm \(\text{tr}((A_n^*A_n)^{1/2})\). We note the submultiplicative inequality \(\|A_nB_n\|_1 \leq \|A_n\|_1\|B_n\|\), or \(\|A_nB_n\|_1 \leq \|B_n\|_2\|A_n\|\), is standard (see \cite{12} p. 454, 7.3.P16). Of course, this also generalizes to \(\|A_nB_n\|_1 \leq \|A_n\|_p\|B_n\|_q\) whenever \(1/p + 1/q = 1\), where \(\|A_n\|_p := \text{tr}((A_n^*A_n)^{p/2})^{1/p}\) denotes the Schatten \(p\)-norm. In addition, we will also be concerned with conditions involving the block size sequence \(\{k_n\}\).

Our most general result (Theorem \(\amalg\)) provides a trace restriction on the rectangular blocks \(A_n, B_n\) when either \(C\) or \(Z\) is in block tridiagonal form, even in the case when both \(C, Z\) are bounded (not necessarily compact) operators.

**Theorem 3.1.** Let \(C, Z \in \mathcal{B}(\mathcal{H})\) and \(T = [C, Z]\). Let \(\langle e_n \rangle\) be any basis in which \(C, Z\) have block matrix form \(3.1\), i.e.,

\[
C = \begin{pmatrix}
C_1 \quad A_1 \quad * \quad \cdots \\
B_1 \quad C_2 \quad A_2 \quad \cdots \\
* \quad B_2 \quad C_3 \quad \cdots \\
\vdots \quad \cdots \quad \cdots \quad \cdots
\end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix}
Z_1 \quad X_1 \quad * \quad \cdots \\
Y_1 \quad Z_2 \quad X_2 \quad \cdots \\
* \quad Y_2 \quad Z_3 \quad \cdots \\
\vdots \quad \cdots \quad \cdots \quad \cdots
\end{pmatrix} \quad (3.2)
\]
both with the central blocks of sizes $k_n \times k_n$ where $(k_n)$ is a sequence of positive integers. Let $Q_n$ denotes the projection onto $V_{j=1}^{k_n} e_j$. If either $C$ or $Z$ is block tridiagonal in (3.2), then

$$\text{tr}(Q_n T Q_n) \leq \|A_n Y_n - X_n B_n\| \leq (\|A_n\|_1 + \|B_n\|_1)\|Z\| \leq k_n (\|A_n\| + \|B_n\|)\|Z\|.$$  \hspace{1cm} (3.3)

Proof. Let $P_n$ denote the projection onto $V_{j=k_{n-1}+1}^{k_n} e_j$, where we set $k_0 = 0$ for convenience. Then clearly $Q_n = P_1 + \cdots + P_n$. Let $T_n$ denote the compression of $T = [C, Z]$ to $P_n H$. Since either $C$ or $Z$ is block tridiagonal, a straightforward computation produces

$$T_1 = A_1 Y_1 - X_1 B_1 + [C_1, Z_1],$$

and for all $n \geq 1$,

$$T_{n+1} = A_{n+1} Y_{n+1} - X_{n+1} B_{n+1} - A_n Y_n + X_n B_n + [C_{n+1}, Z_{n+1}].$$

Since $[C_n, Z_n]$ is a commutator of finite matrices, it has trace zero, and therefore

$$\text{tr}(T_1) = \text{tr}(A_1 Y_1 - X_1 B_1),$$

and for all $n \geq 1$,

$$\text{tr}(T_{n+1}) = \text{tr}(A_{n+1} Y_{n+1} - X_{n+1} B_{n+1} - A_n Y_n + X_n B_n).$$

Consequently, the following sum is telescoping, and hence

$$\text{tr}(Q_n T Q_n) = \sum_{j=1}^{n} \text{tr}(T_j) = \text{tr}(A_n Y_n - X_n B_n).$$  \hspace{1cm} (3.4)

Therefore,

$$|\text{tr}(Q_n T Q_n)| = |\text{tr}(A_n Y_n - X_n B_n)|$$

$$\leq \|A_n Y_n - X_n B_n\|_1$$

$$\leq \|A_n\|_1 \|Y_n\| + \|B_n\|_1 \|X_n\|$$

$$\leq (\|A_n\|_1 + \|B_n\|_1)\|Z\|$$

$$\leq k_n (\|A_n\| + \|B_n\|)\|Z\|. \hspace{1cm} \square$$

Remark. Other choices of Hölder’s inequality yield similar information concerning Schatten $p$-norms which we will not pursue here.

When $T = [C, Z]$ is either trace-class or positive, the sequence $\text{tr}(Q_n T Q_n)$ appearing in (3.3) has a (possibly infinite) limit, which quickly leads to the following corollary. Here, additional information can be extracted when $Z$ is compact.

Corollary 3.2. Let $C, Z \in \mathcal{B}(H)$ and $T = [C, Z]$, and let $\langle e_n \rangle$, $\langle k_n \rangle$ and $\langle Q_n \rangle$ be as in Theorem 3.1, with either $C$ or $Z$ in block tridiagonal form. If $T \neq 0$ is trace-class or positive, then

$$0 < \frac{|\text{tr}(T)|}{\|Z\|} \leq \liminf_{n \to \infty} (\|A_n\|_1 + \|B_n\|_1) \leq \liminf_{n \to \infty} k_n (\|A_n\| + \|B_n\|).$$  \hspace{1cm} (3.5)

Moreover, if $Z$ is compact, then these limits inferior must be fact be infinite.

Proof. To prove (3.5), apply Theorem 3.1, divide by $\|Z\|$ throughout in (3.3) and take the limit inferior. Meanwhile, note that since $T \neq 0$ is trace-class or positive, then the trace is well-defined (including the trace being infinite), and hence

$$0 < \frac{|\text{tr}(T)|}{\|Z\|} = \lim_{n \to \infty} \frac{|\text{tr}(Q_n T Q_n)|}{\|Z\|} = \liminf_{n \to \infty} \frac{|\text{tr}(Q_n T Q_n)|}{\|Z\|}.$$

The claim concerning compact $Z$ takes slightly more work. Note that

$$|\text{tr}(Q_n T Q_n)| \leq \|A_n Y_n - X_n B_n\|_1 \leq \|A_n\|_1 \|Y_n\| + \|B_n\|_1 \|X_n\| \leq (\|A_n\|_1 + \|B_n\|_1)(\|X_n\| + \|Y_n\|),$$
Corollary 3.3. Let $\langle n \rangle$ noting that the left-hand side approaches infinity proves the claim. Therefore, dividing by $\|X_n\| + \|Y_n\|$ in the previous display yields

$$\frac{|\text{tr}(Q_nTQ_n)|}{\|X_n\| + \|Y_n\|} \leq \|A_n\|_1 + \|B_n\|_1.$$

Notice that $\|X_n\|, \|Y_n\| \to 0$ since $Z$ is compact. Taking the limit inferior in the previous display and noting that the left-hand side approaches infinity proves the claim. 

Of course, in the case of interest when both $C, Z$ are compact, this becomes:

Corollary 3.3. Let $C, Z \in \mathcal{K}(\mathcal{H})$ and $T = [C, Z]$, and let $\langle e_n \rangle, \langle k_n \rangle$ and $\langle Q_n \rangle$ be as in Theorem 3.1 with either $C$ or $Z$ in block tridiagonal form. If $T \neq 0$ is trace-class or positive, then

$$\lim_{n \to \infty} k_n(\|A_n\| + \|B_n\|) = \infty = \lim_{n \to \infty} k_n(\|X_n\| + \|Y_n\|)$$

Moreover, the same holds for arbitrary $T$ whenever $\liminf_{n \to \infty} |\text{tr}(Q_nTQ_n)| > 0$.

The next theorem presents an obstruction (explicitly realized in Example 3.6 and the proof of Lemma 3.8) to obtaining a positive diagonal operator as a commutator of a compact operator and a bounded operator, at least one of which is block tridiagonal.

Theorem 3.4. Let $\langle k_n \rangle$ be a sequence of positive integers and $\langle d_n \rangle$ be a nonnegative sequence for which

$$\limsup_{n \to \infty} \frac{1}{k_n} (d_1 + \cdots + d_{k_1 + \cdots + k_n}) > 0,$$

and let $D = \text{diag}(\langle d_n \rangle)$ relative to the basis $\langle e_n \rangle$. Then $D$ is not a commutator of a compact operator $C$ and a bounded operator $Z$, both in block matrix form (3.1) relative to this basis with central block sizes $k_n \times k_n$, where at least one of $C, Z$ is block tridiagonal.

Proof. Suppose that $D$ is a commutator of operators $C, Z$ of the specified form and at least one of these is in block tridiagonal form. We will show that $C$ cannot be compact.

Let $Q_n$ be the projection onto $\bigvee_{j=1}^{k_1 + \cdots + k_n} e_j$. Then by Theorem 3.1

$$d_1 + \cdots + d_{k_1 + \cdots + k_n} = \text{tr}(Q_nDQ_n) \leq k_n(\|A_n\| + \|B_n\|)\|Z\|.$$ 

Dividing by $k_n\|Z\|$ and taking the limit superior, by hypothesis we obtain

$$0 < \limsup_{n \to \infty} \frac{1}{k_n\|Z\|} (d_1 + \cdots + d_{k_1 + \cdots + k_n}) \leq \limsup_{n \to \infty}(\|A_n\| + \|B_n\|).$$

Therefore at least one of $\|A_n\|, \|B_n\|$ does not converge to zero, and hence $C$ cannot be compact. 

Remark. Because it is always possible to rearrange $\langle d_n \rangle$ when $\lim d_n = 0$ so that the limit superior in (3.6) is 0, it becomes clear that Theorem 3.4 is dependent on the order of $\langle d_n \rangle$, even for a fixed sequence of block sizes $(k_n)$.

It is also useful to know that condition (3.6) holds for some rearrangement of $\langle d_n \rangle$ if and only if it holds for its decreasing rearrangement. Indeed, if $\langle d_n^\alpha \rangle$ denotes the decreasing rearrangement of $\langle d_n \rangle$ then

$$\frac{1}{k_n} (d_1 + \cdots + d_{k_1 + \cdots + k_n}) \leq \frac{1}{k_n^\alpha} \left( d_1^\alpha + \cdots + d_{k_1 + \cdots + k_n}^\alpha \right).$$

The previous theorem has the following immediate corollary concerning commutators obtained via the Anderson model detailed in (2.1)–(2.2).
Corollary 3.5. Suppose that \( \langle d_n \rangle \) is a sequence of strictly positive numbers such that
\[
\limsup_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} d_k > 0,
\]
and \( D = \text{diag}(d_n) \). Then \( D \) cannot be obtained using the Anderson model. That is to say, \( D \) is not a commutator of a compact operator \( C \) and a bounded operator \( Z \) both of the form (2.1)–(2.2) with central blocks sizes \( n \times n \). The same holds even more generally when \( A_n, B_n, X_n, \) and \( Y_n \) are possibly full rectangular matrices, or even if the diagonal blocks of \( C \) are nonzero.

Example 3.6. There are many sequences satisfying (3.7), for instance, \( d_n = 1/\sqrt{n} \) (whose partial sums satisfy \( \sum_{j=1}^{n(n+1)/2} d_j = \Theta(n) \) in Bachmann–Landau notation, i.e., \( f(n) = \Theta(g(n)) \) means \( f(n) = O(g(n)) \) and \( g(n) = O(f(n)) \)) is an example where Anderson’s model fails by Corollary 3.5. In Theorem 3.10, we will see that this is a consequence of the arithmetic (hence subexponential) growth of the block sizes in Anderson’s model.

Remark 3.7. We note that when the block sizes \( \langle k_n \rangle \) grow exponentially (i.e., \( k_n \geq \Theta(\rho^n) \) for some \( \rho > 1 \)), then the sequence \( \langle (k_1 + \cdots + k_n)/k_n \rangle \) is bounded above. Then, for any nonnegative sequence \( \langle d_n \rangle \) converging to zero, it is well known that its Cesàro means tend to zero, and hence
\[
\limsup_{n \to \infty} \frac{1}{k_n} (d_1 + \cdots + d_{k_1 + \cdots + k_n}) = \limsup_{n \to \infty} \frac{k_1 + \cdots + k_n}{k_n} \frac{d_1 + \cdots + d_{k_1 + \cdots + k_n}}{k_1 + \cdots + k_n} = \limsup_{n \to \infty} \frac{k_1 + \cdots + k_n}{k_n} \limsup_{n \to \infty} \frac{d_1 + \cdots + d_{k_1 + \cdots + k_n}}{k_1 + \cdots + k_n} = 0,
\]
since \( (k_1 + \cdots + k_n)/k_n \) is bounded above.

The next two lemmas show that for any sequence \( \langle k_n \rangle \) which does not grow sufficiently rapidly, there is a strictly positive nonincreasing sequence \( \langle d_n \rangle \) satisfying condition (3.7). Moreover, this growth condition implies the sequence \( \langle k_n \rangle \) grows at least exponentially.

Lemma 3.8. Let \( \langle k_n \rangle_{n=1}^{\infty} \) be a sequence of positive integers, and let \( \langle s_n \rangle_{n=1}^{\infty} \) denote its sequence of partial sums. If for every nonincreasing positive sequence \( \langle d_n \rangle_{n=1}^{\infty} \) converging to zero one has
\[
\limsup_{n \to \infty} \frac{1}{k_n} \sum_{j=1}^{s_n} d_j < \infty,
\]
then \( \liminf_{n \to \infty} k_n/s_n > 0 \).

Proof. We will prove the contrapositive. Suppose that \( \liminf_{n \to \infty} k_n/s_n = 0 \). Then \( \limsup_{n \to \infty} s_n/k_n = \infty \). So there is a strictly increasing sequence of positive integers \( \langle n_l \rangle_{l=1}^{\infty} \) such that for all \( l \geq 1 \), \( s_{n_l}/k_{n_l} > l \).

For convenience of notation, set \( n_0 := 0 \) and \( s_0 := 0 \). Then define a positive nonincreasing sequence \( \langle d_j \rangle_{j=1}^{\infty} \) converging to zero by \( d_j := (\log(l+1))/l \) for \( s_{n_l-1} < j \leq s_{n_l} \). Then we clearly have for all \( l \),
\[
\frac{1}{s_{n_l}} \sum_{j=1}^{s_{n_l}} d_j \geq \frac{\log(l+1)}{l}.
\]
Consequently,
\[
\frac{1}{k_{n_l}} \sum_{j=1}^{s_{n_l}} d_j = \frac{s_{n_l}}{k_{n_l}} \cdot \frac{1}{s_{n_l}} \sum_{j=1}^{s_{n_l}} d_j \geq l \cdot \frac{\log(l+1)}{l} = \log(l+1).
\]
Hence \( \limsup_{n \to \infty} \frac{1}{k_n} \sum_{j=1}^{s_n} d_j = \infty \) \( \square \)

Lemma 3.9. Let \( \langle k_n \rangle_{n=1}^{\infty} \) be a sequence of positive integers, and let \( \langle s_n \rangle_{n=1}^{\infty} \) denote its sequence of partial sums. If \( \liminf_{n \to \infty} k_n/s_n > 0 \), then \( k_n = \Omega(\rho^n) \) for some \( \rho > 1 \) (Bachmann–Landau notation, \( f(n) = \Omega(g(n)) \) is \( g(n) = O(f(n)) \)). In fact, for any \( 0 < \delta < \liminf_{n \to \infty} k_n/s_n \), we may choose \( \rho = 1 + \delta \).
Proof. Take any $0 < \delta < \liminf_{n \to \infty} k_n/s_n$, and let $\rho = 1 + \delta$. Then there is some $N \geq 1$ such that $k_n/s_n > \delta$ for all $n \geq N$. Then for any $n \geq N$,

$$s_{n+1} = k_{n+1} + s_n > \delta s_{n+1} + s_n > \rho s_n.$$ 

Now a simple induction shows, for all $n \geq N$, $s_n \geq \rho^{n-N}s_N$, and hence $k_n \geq \delta\rho^{n-N}s_N$, so $k_n = \Omega(\rho^n)$. \hfill \square

**Theorem 3.10.** Let $(k_n)$ be a sequence of positive integers. Suppose that for every strictly positive non-increasing sequence $(d_n)$ converging to zero, $D = \text{diag}(d_n)$ (relative to a basis $(e_n)_{n=1}^\infty$) is a commutator $[C, Z]$ where $C$ is compact and either $C$ or $Z$ has block tridiagonal form (relative to $(e_n)_{n=1}^\infty$) with central block sizes $k_n$. Then $(k_n)$ has at least exponential growth: $k_n = \Omega(\rho^n)$ for some $\rho > 1$.

Proof. Suppose that $(k_n)$ is such a sequence as specified in the theorem. Then by hypothesis, for any nonincreasing positive sequence $(d_n)$ converging to zero, diag$(d_n) = [C, Z]$ for some compact operator $C$ such that either $C$ or $Z$ is block tridiagonal. Then the contrapositive of Theorem 3.4 guarantees

$$\limsup_{n \to \infty} \frac{1}{k_n} (d_1 + \cdots + d_{k_1 + \cdots + k_n}) = 0.$$ 

Since this holds for all positive nonincreasing sequences converging to zero, Lemmas 3.8 and 3.9 prove the desired claim. \hfill \square

The exponential growth requirement from Theorem 3.10 together with the disappearance of the obstruction \ref{Obs} when $(k_n)$ grows exponentially (see Remark \ref{remark}.) suggest the possibility that every strictly positive diagonal operator is a commutator of compact operators via block tridiagonal operators with exponentially growing block sizes. We will investigate this idea further in Section 4, but given the relative success of Anderson’s model, it is natural to wonder whether this is possible with an exponential version of Anderson’s model (EAM) displayed in (\ref{3.8}) below. Namely, let $(k_n)$ be a sequence of positive integers satisfying $k_{n+1} \geq qk_n$ for some $q > 1$ (which includes the case equivalent to $a\rho^n$ for integers $a > 0, \rho > 1$), and let

$$A_n = \begin{pmatrix} a_{1,n} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & a_{2,n} & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & a_{k_n,n} & 0 & \cdots & 0 \end{pmatrix}, \quad X_n = \begin{pmatrix} 0 & x_{1,n} & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & 0 & x_{2,n} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & x_{k_n,n} & 0 & \cdots & 0 \end{pmatrix}, \quad (3.8)$$

$$B_n = \begin{pmatrix} 0 & \cdots & 0 & 0 \\ -b_{1,n} & 0 & \cdots & 0 \\ 0 & -b_{2,n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & -b_{k_n,n} \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}, \quad Y_n = \begin{pmatrix} y_{1,n} & 0 & \cdots & 0 \\ 0 & y_{2,n} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & y_{k_n,n} \\ 0 & \cdots & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix},$$

where these matrices are of the sizes $k_n \times k_{n+1}$ ($A_n$ and $X_n$) and $k_{n+1} \times k_n$ ($B_n$ and $Y_n$). The model (EAM) is obtained by defining $C, Z$ as in (\ref{2.1}), but with the central blocks of size $k_n \times k_n$. It is natural to have

$$\frac{s_n - s_{n-1}}{n - (n-1)} = k_n > \delta s_n.$$

Note that this is essentially a discrete differential inequality since

$$\frac{s_n - s_{n-1}}{n - (n-1)} = k_n > \delta s_n.$$
to ask whether the commutator $CZ - ZC$ can yield every compact diagonal operator. We show that the answer remains: No!

**Theorem 3.11.** If the (EAM) yields every compact diagonal operator as a commutator of compact operators, then so also does the (AM).

Corollary 3.5 and Example 3.6 show that the (AM) cannot accomplish this goal so it follows by contrapositive that neither can the (EAM).

**Corollary 3.12.** Not all strictly positive diagonal operators can be represented as commutators of compact operators using (EAM).

**Proof of Theorem 3.11.** Suppose that $D' = \text{diag}(D'_n)$ is a compact operator, with each block $D'_{n}$ being a diagonal $n \times n$ matrix. For each $n \in \mathbb{N}$, let $D_n$ be a $k_n \times k_n$ diagonal matrix such that the upper left $n \times n$ block is $D'_n$, i.e.,

$$D_n = \begin{pmatrix} D'_n & 0 \\ 0 & \ast \end{pmatrix}.$$  

Of course, if $k_n = n$ then $D_n = D'_n$, and we notice that this can happen only for $k_1 = 1$. Since (EAM) by hypothesis can represent $D = \text{diag}(D_n)$ as a commutator of compact operators $C, Z$ as in (2.3), but with central blocks of size $k_n \times k_n$, and with blocks satisfying for $n \geq 1$,

$$A_n X_{n+1} = X_n A_{n+1} \quad \text{and} \quad B_{n+1} Y_n = Y_{n+1} B_n,$$  

$$D_n = A_n Y_n - X_n B_n + B_{n-1} X_{n-1} - Y_{n-1} A_{n-1},$$  

(3.9) taking $A_0 = B_0 = X_0 = Y_0 = 0$. Furthermore, the blocks in (EAM) are given by (3.8) and it is not hard to see that each can be written in the form

$$A_n = \begin{pmatrix} A'_n & 0 \\ \ast & \ast \end{pmatrix}, \quad X_n = \begin{pmatrix} X'_n & 0 \\ 0 & \ast \end{pmatrix}, \quad B_n = \begin{pmatrix} B'_n & 0 \\ 0 & \ast \end{pmatrix}, \quad Y_n = \begin{pmatrix} Y'_n & 0 \\ 0 & \ast \end{pmatrix},$$  

(3.11) where the blocks $A'_n, X'_n, B'_n, Y'_n$ are given by (2.2). (Once again, if $k_1 = 1$, then $A_1 = A'_1$.) Substituting (3.11) into (3.9)–(3.10) we obtain

$$\begin{pmatrix} A'_n & X'_n \\ \ast & \ast \end{pmatrix} = \begin{pmatrix} X'_n & A'_{n+1} \\ \ast & \ast \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} B'_{n+1} & Y'_n \\ \ast & \ast \end{pmatrix} = \begin{pmatrix} Y'_n & B'_n \\ \ast & \ast \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} D'_n & 0 \\ 0 & \ast \end{pmatrix} = \begin{pmatrix} A'_n & Y'_n \\ \ast & \ast \end{pmatrix} - \begin{pmatrix} X'_n & B'_n \\ \ast & \ast \end{pmatrix} + \begin{pmatrix} B'_{n-1} & X'_{n-1} \\ \ast & \ast \end{pmatrix} - \begin{pmatrix} Y'_{n-1} & A'_n \\ \ast & \ast \end{pmatrix}. \quad \text{(3.12)}$$

The only calculation that may not be immediately obvious concerns the product $Y_{n-1} A_{n-1}$ which is a diagonal matrix, and its upper left corner is indeed $Y'_{n-1} A'_{n-1}$. It follows from (3.12) that the smaller blocks $A'_{n'}, X'_{n'}, B'_{n'}, Y'_{n'}$ satisfy conditions (3.9)–(3.10). Consequently, $D' = [C', Z']$, where $C', Z'$ are of the form (2.1) but with blocks $A'_{n'}, B'_{n'}, X'_{n'}, Y'_{n'}$. In other words $D'$ is a commutator of compact operators given by (AM).

4. **Universal Block Tridiagonalization and the Generalized Anderson Model (GAM)**

In this section we will consider the problem of general matrix sparsification under a special unitary equivalence. One approach to the general problem for arbitrary $B(\mathcal{H})$-operators is staircase forms given in [22], which was an attempt to extend to all operators the well-known tridiagonalization of selfadjoint operators that possess a cyclic vector. This was essential to the solution in [22] to [19] Problems 3 and 3’ and led to [18], a precursor to this paper. Some of the material presented in this section appears in [18] but we include it here for completeness, enhanced and somewhat reformulated but with less detail (Theorem 4.3, Remark 4.5).

The following general staircase form (4.1) is obtained in [22] by considering the free semigroup on the two generators $C, C^*$ with any basis $\langle e_n \rangle$ to generate a new basis via applying Gram–Schmidt to $e_1, C e_1,$
This list consists of \( W(C, C^*) \) (all words in \( C, C^* \), including the empty word \( I \)) applied to all elements of \( \langle e_n \rangle \) and arranged in this special order: first list \( e_1 \) followed by \( C, C^* \) applied to \( e_1 \) to obtain the first three vectors; then list \( e_2 \), followed by \( C, C^* \) applied to the second vector in this list, namely \( C e_1 \); and so on. By a \textit{staircase form}, we mean a matrix and two strictly increasing sequences of positive integers \( \langle r_1(n) \rangle, \langle r_2(n) \rangle \), called the \textit{column and row support lengths}, for which the matrix entries are zero in the positions \( (i, j) \) whenever \( i > r_1(j) \) or \( j > r_2(i) \).

**Theorem 4.1** ([22, Theorem 2–Corollary 3, combined]). Let \( C \in B(H) \) be arbitrary and let \( \langle e_n \rangle \) be any orthonormal basis of \( H \). There is an orthonormal basis \( \langle f_n \rangle \) is derived from \( \langle e_n \rangle \) (via the method in the above paragraph) such that \( f_1 = e_1 \) and \( e_n \in \vee_{k=1}^{3n} f_k \) for all \( n \in \mathbb{N} \), and moreover, the matrix of \( C \) with respect to \( \langle f_n \rangle \) takes the staircase form

\[
C = \begin{pmatrix}
* & * & * & 0 & \cdots \\
* & * & * & * & \cdots \\
* & * & \cdots \\
0 & * \\
0 & * \\
\vdots & 0 
\end{pmatrix},
\]  

(4.1)

where this matrix has row and column support lengths \( \langle 3n \rangle_{n=1}^\infty \).

In addition, if \( S_1, \ldots, S_N \) is any finite collection of selfadjoint operators, then there is an orthonormal basis \( \langle f_n \rangle \) for which \( f_1 = e_1 \) and \( e_n \in \vee_{k=1}^{r_f(n)} f_k \), and with respect to which \( S_k \) takes the form \( \text{(4.1)} \) with \( \langle 3n \rangle_{n=1}^\infty \) replaced by \( \langle (N + 1)n \rangle_{n=1}^\infty \).

When \( S_1, \ldots, S_N \) are not necessarily selfadjoint, the row and column support lengths are given by \( \langle (2N + 1)n \rangle_{n=1}^\infty \), and the inclusion relationships are \( e_n \in \vee_{k=1}^{r_f(n)} f_k \).

**Definition 4.2.** We call the \(*\)-entries in \( \text{(4.1)} \) the support entries. Albeit some can also be zero as we shall see in Theorems 4.9 and 4.11. So for instance, for \( N = 1 \) we obtained support row and column sizes \( \langle 3n \rangle \), which technically have fewer support entries because of the initial zeros (starting for instance with the 4th row and column which has 11 asterisks and not 12), but for simplicity we don’t mention those.

Theorem 4.1 gives a striking universal staircase form (with step sequence \( \langle 3n \rangle \)) for an arbitrary operator and universal simultaneous staircase forms with larger stairs for arbitrary finite collections. One can slightly strengthen Theorem 4.1 to have column lengths \( \langle 3n - 1 \rangle \) (see [13, Remark 20.3]).

The staircase form \( \text{(4.1)} \) will allow us to represent a matrix in universal block tridiagonal form as in \( \text{(4.2)} \) below:

\[
C = \begin{pmatrix}
C_1 & A_1 & 0 & \cdots \\
B_1 & C_2 & A_2 & \cdots \\
0 & B_2 & C_3 & \cdots \\
\vdots & \ddots & \ddots & \ddots 
\end{pmatrix},
\]

(4.2)

The following theorem gives canonical dimensions for the blocks in \( \text{(4.2)} \) in order that they cover all the support entries. And since there will be no change of basis, we retain the Theorem 4.1 condition that each \( e_n \in \vee_{k=1}^{3n} f_k \).

\footnote{When the list is linearly dependent, an induction argument is necessary.}
Theorem 4.3. For all $C \in \mathcal{B}(\mathcal{H})$, the block tridiagonal partition of the matrix of $C$ induced by (4.1) (i.e., in the derived basis $(f_n)$ from Theorem 4.1) is given by (4.2), where the central blocks $C_n$ have block size sequence $(k_n)$ given by $k_1 = 1$ and $k_n = 2(3^{n-2})$ for $n > 1$.

More generally, the necessary and sufficient conditions that (4.2) cover the (4.1) staircase support entries are: $k_1$ is chosen arbitrarily and $k_{n+1} \geq 2(k_1 + k_2 + \cdots + k_n)$.

Proof. Let $(s_n)$ denote the sequence of partial sums of the block size sequence $(k_n)$. Note that an entry in the $(i,j)$-position of $C$ is a support entry with respect to the basis $(f_n)$ if and only if $i \leq 3j$ and $j \leq 3i$.

The recursive condition $k_{n+1} \geq 2s_n$ is clearly necessary since the $(s_n, 3s_n)$-position is a support entry of $C$, and hence $3s_n \leq s_{n+1}$ if the block tridiagonal form covers this support entry.

For the sufficiency, suppose that $k_{n+1} \geq 2s_n$ for all $n \in \mathbb{N}$, so then $3s_n \leq s_{n+1}$. Suppose the $(i,j)$-position is a support entry of $C$. Then there is some $n \in \mathbb{N}$ for which $s_{n-1} < i \leq s_n$ (where we let $s_0 = 0$). Then $j \leq 3s_n \leq s_{n+1}$. Similarly, for some $m \in \mathbb{N}$, $s_{m-1} < j \leq s_m$ and hence $i \leq s_{m+1}$. Hence the $(i,j)$-position is covered by the block tridiagonal form (4.2).

Finally, $k_n = 2(3^{n-2})$ for $n > 1$ is obtained by letting $s_1 = 1$ and using equality in the recursive formula.

A similar result holds for more than one matrix. For example, for 3 matrices we have the following consequence of the staircase form for the case $N = 3$ of Theorem 4.1. This is explained further in Remark 4.5. And then a modified argument from the above paragraph shows the blocks cover the support entries of the larger staircase forms.

Theorem 4.4. For any $N$ operators in $\mathcal{B}(\mathcal{H})$, the block tridiagonal partition of their matrices induced by the analog of (4.1) for $N$ (i.e., in the new basis $(f_n)$) is given by (4.2). For each of them simultaneously the central blocks have block size sequence $(k_n)$ given by $k_1 = 1$ and $k_n = 2N((2N + 1)^{n-2})$ for $n > 1$.

In particular, for 3 operators $k_n = 6(7^{n-2})$, and for 2 operators $k_n = 4(5^{n-2})$ for $n > 1$.

A brief explanation for this theorem is:

Remark 4.5. We mention that in the case of more than one operator, the orthonormal basis $(f_n)$ of $\mathcal{H}$ — with respect to which their matrices are in the block tridiagonal form of Theorem 4.1 — can be obtained in the following manner as in [22] Theorem 2—Corollary 3, combined, at least when the list of vectors below is linearly independent. First we decompose each of the operators as $T = T_1 + iT_2$, with $T_1$ and $T_2$ selfadjoint, and then we create a list of vectors. For instance, in the case of 3 operators $C = C_1 + iC_2$, $Z = Z_1 + iZ_2$, and $D = D_1 + iD_2$, we use

$$
e_1, C_1e_1, C_2e_1, Z_1e_1, Z_2e_1, D_1e_1, D_2e_1,$$
$$e_2, C_1g_2, C_2g_2, Z_1g_2, Z_2g_2, D_1g_2, D_2g_2,$$
$$e_3, C_1g_3, C_2g_3, Z_1g_3, Z_2g_3, D_1g_3, D_2g_3,$$

etc.,

where $g_2$ is the second vector, namely $C_1e_1$, $g_3$ is the third vector, namely $C_2e_1$, etc. Finally, Gram–Schmidt is applied to this list. This yields the staircase matrix analog of (4.1) but with support row and column sizes $(7n)$, replacing $(3n)$. After Gram–Schmidt to obtain the analog staircase form, we then partition the staircase matrix into central, upper and lower blocks similarly as in the proof of Theorem 4.1 based on the staircase matrix with row and column sizes $(7n)$. And similarly for the case $N = 2$ with analog staircase matrix row and column sizes $(5n)$.

Theorems 4.3 and 4.4 show that, when considering the equation $CZ - ZC = D$, there is no loss of generality in assuming that any two or all three operators $C, Z, D$ are in the block tridiagonal form (4.2). We formulate this equivalence and leave the proof to the reader.

Theorem 4.6. The following are equivalent:

(i) $T$ is a commutator of compact operators.
(ii) $T$ is a block tridiagonal operator of the form (4.2) relative to a basis $\langle f_n \rangle$, derived via Theorems 4.1, 4.3, and 4.4 from any basis $\langle e_n \rangle$, and $T$ is a commutator of compact matrices of the same block tridiagonal form relative to that same basis (i.e., all with central block sizes $6(7^{n-2})$ after $k_1 = 1$).

For diagonal matrices, the following are equivalent:

(iii) The diagonal matrix $D$ (relative to a basis $\langle e_n \rangle$) is a commutator of compact operators.

(iv) Relative to the basis $\langle f_n \rangle$, derived via Theorems 4.1, 4.3, and 4.4 from $\langle e_n \rangle$, $D$ is a block 5-diagonal operator of the form Equation (4.2), and $D$ is a commutator of compact operators of the same block tridiagonal form relative to that same basis, all with central block sizes $4(5^{n-2})$ after $k_1 = 1$.

**Remark 4.7.** For positive operators, Theorem 4.6 would include the additional sometimes useful inequalities that allow one to regulate the growth rate of the partial traces by selectively choosing various $T$ or $D$ targets. For $T$ or $D = CZ − ZC$, one has the partial traces relative to the bases $\langle e_n \rangle, \langle f_n \rangle$ related via $\sum_i [n/7](Te_i, e_i) ≤ \sum_i [n/5](De_i, e_i) ≤ \sum_i [n/5](De_i, e_i) ≤ \sum_i [n/5](De_i, e_i)$. These can be deduced from the $e_n, f_n$ relations in Theorem 4.1 which remain in effect for Theorems 4.3 and 4.4 applied as earlier in this paper to obtain bounds for trace computations of $D_1 + \cdots + D_n$.

Moreover, we have the following quantitative contraints on block norms: (4.4) and (4.5) below, evolving from (3.3), on the speed of convergence to 0 of

$$\|A_n\|\|Y_n\| + \|X_n\|\|B_n\| ≤ (\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|)$$

after block tridiagonalizing $C, Z$ in the basis $\langle f_n \rangle$.

Let $D = \text{diag}(d_n)$, $d_n \downarrow 0$ be a diagonal operator in the basis $\langle e_n \rangle$. Theorems 4.3 and 4.4 allow us to derive a new basis $\langle f_n \rangle$ with respect to which $C, Z$ are simultaneously in block tridiagonal form with central block sizes $k_n = 4 \cdot 5^{n-2}, k_1 = 1$:

$$C = \begin{pmatrix} C_1 & A_1 & 0 & \cdots \\ B_1 & C_2 & A_2 & \cdots \\ 0 & B_2 & C_3 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix} \quad \text{and} \quad Z = \begin{pmatrix} Z_1 & X_1 & 0 & \cdots \\ Y_1 & Z_2 & X_2 & \cdots \\ 0 & Y_2 & Z_3 & \cdots \\ \vdots & \ddots & \ddots & \ddots \end{pmatrix}. \tag{4.3}$$

The diagonal entries $\langle d'_n \rangle$ of $D$ in the basis $\langle f_n \rangle$ are given by $d'_n = (Df_n, f_n)$. Let $D_n$ be the $n$-th diagonal block of $D = CZ − ZC$ with respect to the block 5-diagonal form on $D$ induced by (4.3). The inclusion relations governing $e_n, f_n$, along with the techniques from Sections 2 and 3 guarantee

$$d_1 + d_2 + \cdots + d_{\lfloor k_1 + \cdots + k_n \rfloor/5} ≤ d'_1 + d'_2 + \cdots + d'_{\lfloor k_1 + \cdots + k_n \rfloor} = \text{tr}(D_1 + D_2 + \cdots + D_n)$$

$$= \text{tr}(A_n Y_n - X_n B_n) ≤ \|A_n Y_n - X_n B_n\|_1$$

(then applying Hölder)

$$≤ k_n(\|A_n\|\|Y_n\| + \|X_n\|\|B_n\|) ≤ k_n(\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|).$$

Similarly, for diagonal $D = \text{diag}(d_n)$, $d_n \downarrow 0$ in the basis $\langle e_n \rangle$, its derived basis $\langle f_n \rangle$ (via Theorem 4.4, $N = 3$) simultaneously puts all three $D, C, Z$ in block tridiagonal form of central block sizes $k_n = 6 \cdot 7^{n-2}, k_1 = 1$. Then

$$d_1 + d_2 + \cdots + d_{\lfloor k_1 + \cdots + k_n \rfloor/7} ≤ d'_1 + d'_2 + \cdots + d'_{\lfloor k_1 + \cdots + k_n \rfloor} = \text{tr}(D_1 + D_2 + \cdots + D_n)$$

$$= \text{tr}(A_n Y_n - X_n B_n) ≤ \|A_n Y_n - X_n B_n\|_1$$

(then applying Hölder)

$$≤ k_n(\|A_n\|\|Y_n\| + \|X_n\|\|B_n\|) ≤ k_n(\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|).$$

From these we obtain the respective quantitative constraints on the speed of convergence to 0 of

$$\|A_n\|\|Y_n\| + \|X_n\|\|B_n\| ≤ (\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|):$$
For $N = 2$, i.e., $C, Z$, we have

$$\frac{1}{k_n} (d_1 + d_2 + \ldots + d_{(k_1 + \ldots + k_n)/5}) \leq \|A_n\|\|Y_n\| + \|X_n\||B_n\| \leq (\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|)$$

(4.4)

and for $N = 3$, i.e., $D, C, Z$, we have

$$\frac{1}{k_n} (d_1 + d_2 + \ldots + d_{(k_1 + \ldots + k_n)/7}) \leq \|A_n\|\|Y_n\| + \|X_n\||B_n\| \leq (\|A_n\| + \|B_n\|)(\|Y_n\| + \|X_n\|).$$

(4.5)

These left-hand ratio sequences (in these exponential cases are related to averages) decreases to 0 as discussed in Remark 4.5. And so this lower bound provides quantitative constraints on the speed of tending to 0 of the right-hand sides, which offers information on the norms of the finite blocks in the block tridiagonalizations of $C, Z$.

**Generalized Anderson’s model (GAM): full finite blocks.** Theorem 4.6(c)–(d) indicates that allowing full block tridiagonal operators (as opposed to the more restricted forms of (EAM) in (3.8)) is in a sense necessary and sufficient to solve the compact commutator problem with one exception. Namely, that the target operator $D$, after the required basis change, is not in block tridiagonal form (though because $C, Z$ are simultaneously in block tridiagonal form with same compatible block sizes, $D$ is in block 5-diagonal form). In contrast, in the beginning in [1] Anderson directly forced a matrix solution for his commutator problem for a simple fixed rank one projection ($p_{11} = 1, p_{ij} = 0$ elsewhere) rather than effecting a change of basis. We do not know whether this strategy always works, so we leave it as one of the essential open questions in the subject:

**Problem 4.8.** Is the change of basis in Theorem 4.4(b) necessary when $T$ is already block tridiagonal relative to the basis $\{e_n\}$? That is, are the conditions in Theorem 4.4 equivalent to: Every compact operator in our block tridiagonal form (4.2) is a commutator of compact operators of the same form relative to the same basis, all 3 with central block sizes $6(7^{n-2})$. This requires solving the algebraic system of equations expanded from display (2.3) to compute commutators $T = C Z - Z C$ of the matrices simultaneously of the form (4.2). Let

$$T = \begin{pmatrix} D_1 & E_1 & 0 & \cdots \\ F_1 & D_2 & E_2 & \cdots \\ 0 & F_2 & D_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \quad C = \begin{pmatrix} C_1 & A_1 & 0 & \cdots \\ B_1 & C_2 & A_2 & \cdots \\ 0 & B_2 & C_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}, \quad \text{and} \quad Z = \begin{pmatrix} Z_1 & X_1 & 0 & \cdots \\ Y_1 & Z_2 & X_2 & \cdots \\ 0 & Y_2 & Z_3 & \cdots \\ \vdots & \vdots & \ddots & \ddots \end{pmatrix}$$

(4.6)

with the central blocks $D_n, C_n, Z_n$ of size $k_n \times k_n$ ($k_n$ as in Theorem 4.4 for $N = 3$). Recall that the target operator $T$ is compact if and only if its three sequences of diagonal (main, upper and lower) finite blocks converge in norm to 0 (see for instance Theorem 6.8 and recall that compact operators are those which take weakly convergent sequences of its Hilbert space vectors to norm convergent sequences). So Anderson’s approach becomes: to solve the Peary–Topping problem it suffices to solve the system of matrix equations for $C, Z$ with their sequences of (main, upper and lower) diagonal blocks also converging in norm to 0. For the record, this expanded algebraic system is:

$$D_1 = A_1 Y_1 - X_1 B_1 + [C_1, Z_1], \quad \text{and for } n \geq 1,$$

$$D_{n+1} = B_n X_n - Y_n A_n + A_{n+1} Y_{n+1} - X_{n+1} B_{n+1} + [C_{n+1}, Z_{n+1}],$$

$$A_n X_{n+1} = X_n A_{n+1} \quad \text{and} \quad B_{n+1} Y_n = Y_{n+1} B_n,$$

$$E_n = C_n X_n - Z_n A_n + A_{n+1} Z_{n+1} - X_{n+1} C_{n+1},$$

$$F_n = B_n Z_n - Y_n C_n + C_{n+1} Y_{n+1} - Z_{n+1} B_n.$$
Observe that even in this generality the important relation in (3.4) and used often above is preserved:
\[ \text{tr}(D_1 + D_2 + \cdots + D_n) = \text{tr}(A_n Y_n - X_n B_n). \]
Also, when \( T \) is simply a diagonal matrix \( D \), importantly this system is much simpler because each \( E_n, F_n \) is the zero matrix of the corresponding size, and all \( D_n \) are square diagonal matrices. Thus this system has a better chance of being solvable or leading to counterexample obstructions.

**Further sparsification for block tridiagonal forms.** It is natural to ask whether Theorem 4.3 can be further improved, i.e., can more zeros be obtained in the blocks. Here, we are attempting to preserve the structure and block sizes as in Theorem 4.3, while ensuring that some additional entries in these blocks further improved, i.e., can more zeros be obtained in the blocks. Here, we are attempting to preserve the structure and block sizes as in Theorem 4.3, while ensuring that some additional entries in these blocks are universally zero. Theorems 4.9 and 4.11, whose proofs can be found in [18], present some progress in this direction.

**Theorem 4.9** ([18], Theorem 20.7). Every block tridiagonal matrix with diagonal block sizes \( \{k_n\} \) non-decreasing is unitarily equivalent to a block tridiagonal matrix with the same block sizes but also with \( A_n \) of the form \( (A'_n \ | \ 0) \) with \( A'_n \) a positive square matrix. Alternatively, the same but with \( B_n \) of the form \( (B'_n \ | \ 0)^T \) with \( B'_n \) a positive square matrix.

The next theorem ([18], Theorem 20.8), is a further substantial sparsification of our canonical block tridiagonal Theorem 4.3 case (i.e., with central block sizes \( k_1 = 1, k_n = 2 \cdot 3^{n-2} \)). However there is a subtle gap in the proof appearing in [18], namely, the linearly dependent case at the end. (See description just below.) So here we give a different proof modeled after the proof of [22], Theorem 2]. It is more complex than [18] Theorem 20.8 so we first give a motivation.

Our study of block tridiagonal forms began with a study of staircase forms (see Theorem 4.1 and preceding paragraph) via the sequence \( e_1, C e_1, C^* e_1, e_2, C^2 e_1, C^* C e_1, e_3, C C^* e_1, C^* e_2, e_4, C e_2, C^* e_2, e_5, \ldots \). This sequence generates the staircase form (4.1) when it is linearly independent, and it takes an induction proof to handle the retention of the form in the linearly dependent case. The core idea of the next theorem is to redistribute the vectors \( \langle e_n \rangle \) from arithmetically distributed (every third step) to geometrically distributed (every \( 3^{n-1}, n \geq 1 \) step). And again the difficult part, and where the error occurred, is in the linearly dependent case, which is naturally circumvented in the current proof.

We require the following simple lemma concerning the Gram–Schmidt process.

**Lemma 4.10.** Suppose that \( \langle g_n \rangle \) is a linearly independent set such that there is an increasing sequence of positive integers \( m_n \) and an operator \( C \) such that
\[ C g_n \in \bigvee_{j=1}^{m_n} g_j \text{ for all } n \in \mathbb{N}. \]
Then if \( \langle f_n \rangle \) is the orthonormal set obtained by applying the Gram–Schmidt process to \( \langle g_n \rangle \), then
\[ C f_n \in \bigvee_{j=1}^{m_n} f_j \text{ for all } n \in \mathbb{N}. \]
Consequently, the matrix representation for \( C \) relative to the orthonormal basis \( \langle f_n \rangle \) has the property that, for the \( n \)-th column, the length of the support entries is at most \( m_n \).

**Proof.** Since \( \langle g_n \rangle \) is a linearly independent set, the Gram–Schmidt process guarantees that \( \bigvee_{j=1}^{n} f_j = \bigvee_{j=1}^{n} g_j \) for all \( n \). Note that for all \( n \in \mathbb{N} \), by hypothesis and the linearity of \( C \),
\[ C \left( \bigvee_{j=1}^{n} g_j \right) = \bigvee_{j=1}^{n} C g_j \leq \bigvee_{j=1}^{m_n} \left( \bigvee_{j=1}^{m_n} g_j \right) = \bigvee_{j=1}^{m_n} g_j, \]
where the latter equality holds since \( m_n \) is an increasing sequence. Consequently,

\[
C f_n \in C \left( \bigvee_{j=1}^{n} f_j \right) = C \left( \bigvee_{j=1}^{m_n} g_j \right) \subseteq \bigvee_{j=1}^{m_n} g_j = \bigvee_{j=1}^{m_n} f_j.
\]

**Theorem 4.11.** Let \( \langle e_k \rangle_{k=0}^m \) be a countable (finite or infinite) collection of strictly increasing functions whose ranges together constitute a partition of \( \mathbb{N} \), and let \( r_0(1) = 1 \). Let \( \langle T_k \rangle_{k=1}^m \in \mathcal{B}(\mathcal{H}) \) and let \( \langle e_n \rangle \) be a spanning set for \( \mathcal{H} \). Consequently, the sequence of column support lengths of \( m \), where the latter equality holds since \( m_n \) is an increasing sequence. Consequently, \( m_n \) is defined previously since \( m_n \) is strictly increasing, \( r_0(1) = 1 \) and the ranges are disjoint, \( l < r_k(l) \).

We construct a linearly independent spanning set \( \langle g_n \rangle \) inductively as follows. Set \( g_1 := e_1 = e_{r_0^{-1}(1)} \).

For \( n > 1 \), there is a unique \( 0 \leq k \leq m \) for which \( n \in r_k(\mathbb{N}) \). If \( k = 0 \), define \( g_n \) to be the first vector in the list:

\[
\{ e_{r_0^{-1}(n)}, e_{r_0^{-1}(n)+1}, \ldots \} \setminus \bigvee_{j=1}^{n-1} g_j.
\]

(4.8)

Otherwise, if \( 1 \leq k \leq m \) define \( g_n \) to be the first vector in the list:

\[
\{ T_k g_{r_k^{-1}(n)}, e_2, e_3, \ldots \} \setminus \bigvee_{j=1}^{n-1} g_j.
\]

(4.9)

Note that \( g_{r_k^{-1}(n)} \) is defined previously since \( r_k^{-1}(n) < r_k(r_k^{-1}(n)) = n \). Moreover, in each case this list is nonempty since \( \langle e_n \rangle \) is a spanning set for \( \mathcal{H} \) and the span \( \bigvee_{j=1}^{n-1} g_j \) is finite dimensional. Also, the sequence so constructed is linearly independent by design.

Therefore, by construction, for each \( n \in \mathbb{N} \), and for all \( 1 \leq k \leq m \)

\[
e_n = e_{r_0^{-1}(r_0(n))} \in \bigvee_{j=1}^{r_0(n)} g_j \quad \text{and} \quad T_k g_n = T_k g_{r_k^{-1}(r_k(n))} \in \bigvee_{j=1}^{r_k(n)} g_j.
\]

The first of these conditions guarantees that \( \langle g_n \rangle \) spans \( \mathcal{H} \) since \( \langle e_n \rangle \) does, and hence applying Gram–Schmidt to \( \langle g_n \rangle \) yields an orthonormal basis \( \langle f_n \rangle \). By Lemma 4.10 and the fact that \( \bigvee_{j=1}^{n} f_j = \bigvee_{j=1}^{n} g_j \) for all \( n \), we may replace \( \langle g_n \rangle \) with \( \langle f_n \rangle \) everywhere in the previous display.

Finally, if each \( T_k \) is upper triangular relative to an orthonormal basis \( \langle e_n \rangle \) one can show that \( g_n = e_n \) for every \( n \), and hence \( f_n = e_n \) also. Indeed, it is a straightforward induction. Note that \( g_2 = e_1 \). Let \( n > 1 \) and suppose that \( g_j = e_j \) for all \( 1 \leq j < n \). When considering the definition of \( g_n \), there are two cases depending on whether \( n \in r_0(\mathbb{N}) \) or \( n \in r_k(\mathbb{N}) \) for some \( 1 \leq k \leq m \). In the former case, \( g_n \) is defined as the first vector in the list \( \{ \}

(4.8)

Note that \( \langle e_n \rangle \) is strictly increasing, \( r_0(1) = 1 \) and the fact that each \( T_k \) is upper triangular with respect to the basis \( \langle e_i \rangle_{i=1}^n \), and since \( r_k^{-1}(n) < n \),

\[
T_k g_{r_k^{-1}(n)} = T_k e_{r_k^{-1}(n)} \subseteq \bigvee_{j=1}^{r_k^{-1}(n)} e_j \subseteq \bigvee_{j=1}^{n-1} e_j = \bigvee_{j=1}^{n-1} g_j.
\]
Hence, this first vector is removed from the list \( \{4.9\} \), so \( g_n = e_n \).

\[ \square \]

\textbf{Remark 4.12.} Note that the previous theorem has a curious consequence. For any countably infinite collection of operators \( \{T_k\}_{k=1}^{\infty} \), there is an orthonormal basis \( \langle f_n \rangle \) relative to which the matrix representation of each \( T_k \) has only one column and support lengths (albeit not the same lengths for each operator). Indeed, simply partition \( \mathbb{N} \) into an infinite collection \( \{A_k\}_{k=0}^{\infty} \) of infinite sets with \( 1 \in A_0 \). Then notice that each set in the partition induces a unique strictly increasing function \( r_k : \mathbb{N} \to \mathbb{N} \) with \( r_k(\mathbb{N}) = A_k \). Finally, apply Theorem 4.11 to the sequence \( \langle T_k \rangle_{k=1}^{\infty} \), where \( T_{2k-1} = T_k \) and \( T_{2k} = T_k^* \).

\textbf{Corollary 4.13.} Let \( r_0, r_1, r_2 : \mathbb{N} \to \mathbb{N} \) be strictly increasing functions whose ranges partition \( \mathbb{N} \), and let \( r_0(1) = 1 \). Let \( T \in \mathcal{B}(\mathcal{H}) \) be arbitrary, and let \( \langle e_n \rangle \) be any spanning set. Then there is an orthonormal basis \( \langle f_n \rangle \) derived from these data relative to which sequences of column and row support lengths of \( T \) are given by \( \langle r_1(n) \rangle, \langle r_2(n) \rangle \), respectively. Moreover, \( e_n \in \bigcup_{j=1}^{r_2(n)} f_j \) for all \( n \in \mathbb{N} \).

\textbf{Proof.} Apply Theorem 4.11 with \( T_1 := T \) and \( T_2 := T^* \).

\[ \square \]

\textbf{Theorem 4.14 (\textsuperscript{[18] Theorem 20.8}).} For arbitrary \( C \in \mathcal{B}(\mathcal{H}) \) and any orthonormal basis \( \langle e_n \rangle \) of \( \mathcal{H} \), there exists an orthonormal basis \( \langle f_n \rangle \) in which \( C \) has a block tridiagonal form as in \( \{4.12\} \) with the block sizes as in Theorem \( \{4.3\} \) (with \( k_1 = 1 \)) and:

\begin{enumerate}[label=(\roman*)]
  \item each block \( B_n = (B_n^r | 0 | 0)^T \) where all three blocks are square and \( B_n^r \) is upper triangular, i.e., \( B_n^r(i, j) = 0 \) if \( i > j \);
  \item each block \( A_n \) is of the form \( (A_n^r | A_n^l | 0) \) where all three blocks are square and \( A_n^r \) is lower triangular, i.e., \( A_n^r(i, j) = 0 \) if \( i < j \);
  \item \( e_1 = f_1 \) and \( e_n \in \bigcup_{k=1}^{r_2(n)} f_k \) for all \( n \in \mathbb{N} \).
\end{enumerate}

Alternatively, \( C \) is unitarily equivalent to another matrix of the form \( \{4.12\} \) with the block sizes as in Theorem \( \{4.3\} \) (with \( k_1 = 1 \)), where each \( A_n = (A_n^r | 0 | 0) \), all three blocks are square, and \( A_n^r \) is lower triangular, while each \( B_n \) has the form \( (B_n^r | B_n^l | 0)^T \), all three blocks are square, and \( B_n^r \) is an upper triangular matrix, and \( \{\text{iii}\} \) holds. (We alert the reader that herein our notation \( k_n \) has been changed from \( n_k \) in \( \{18\} \).

\textbf{Proof.} We apply Theorem 4.11 with \( T_1 := C \), \( T_2 := C^* \), \( r_0(n) := 3^n \) for \( n > 1 \), and \( r_1(1) = 2 \) and for \( n > 1 \),

\[ r_1(n) := 3^{\lceil \log_3 n \rceil} + n - 3^{\lceil \log_3 n \rceil} - 1, \]

and \( r_2 : \mathbb{N} \to \mathbb{N} \) is the unique strictly increasing function whose range is \( \mathbb{N} \setminus (r_0(\mathbb{N}) \cup r_1(\mathbb{N})) \). In particular, the ranges of these functions are:

\[
\begin{align*}
  r_0(\mathbb{N}) &= \{1\} \cup \{3^k \mid k > 1\} \\
  r_1(\mathbb{N}) &= \{2\} \cup \{3^k + j \mid k \geq 1, 1 \leq j \leq 2 \cdot 3^{k-1}\} \\
  r_2(\mathbb{N}) &= \{3\} \cup \{3^k + j \mid k \geq 1, 2 \cdot 3^{k-1} < j < 2 \cdot 3^k\}
\end{align*}
\]

Theorem 4.11 guarantees that the sequences of column and row (column of \( C^* \)) support lengths of \( C \) are given by \( \langle r_1(n) \rangle_{n=1}^{\infty} \) and \( \langle r_2(n) \rangle_{n=1}^{\infty} \), respectively. Therefore, since \( r_1(1) = 2 \) and \( r_2(1) = 3 \), this yields \( C_1, A_1, B_1 \). Moreover, the column lengths being given by \( r_1(n) \) precisely guarantees that each block \( B_k \) (thinking of \( k = \lceil \log_3 n \rceil + 1 \)) has the desired form with \( B_k^r \) upper triangular.

Since the row support lengths for rows 2 and 3 are \( r_2(2) = 6 \) and \( r_2(3) = 7 \), this guarantees \( A_2 \) has the desired form. Each block of entries \( \{3^k + j \mid 2 \cdot 3^{k-1} < j < 2 \cdot 3^k\} \) in \( r_2(\mathbb{N}) \) has size \( 4 \cdot 3^{k-1} - 1 \). Therefore, if \( r_2(n) = 3^k + j \) for some \( k' > 1 \) and \( 2 \cdot 3^{k-1} < j < 2 \cdot 3^{k'-1} \), then because \( r_2 \) is strictly increasing, we
must have
\[ n = 1 + \sum_{i=1}^{k'-1} (4 \cdot 3^{i-1} - 1) + (j - 2 \cdot 3^{k'-1}) = j - k'. \]

Consequently, if also \(3^{k-1} < n \leq 3^k\) (for \(k > 1\)), then we have the inequality
\[ 2 \cdot 3^{k-1} - k' < j - k' = n \leq 3^k, \]
which necessitates \(k' \leq k\). Therefore,
\[ r_2(n) = 3^{k'} + j = 3^k + n + k' \leq 3^k + n + 3^{k-1} = 3^k + (2 \cdot 3^{k-1}) + (n - 3^{k-1}). \]

Since \(k = \lceil \log_3 n \rceil\), and \(r_2(n)\) denotes the row support length, we see that \(A_{k+1}\) has the desired form. \(\square\)

Remark 4.15. In the case where, at every stage, the initial vectors in the list obtained in the proof of Theorem 4.11 are linearly independent from what came before, the construction leads to the following list of vectors:
\[ \langle T_n g_n^{-1}(n) \rangle_{n=1}^{\infty}, \]
where \(\ast\) denotes the unique \(0 \leq k \leq m\) for which \(n \in r_k(\mathbb{N})\), and for convenience we define \(T_0 g_0^{-1}(n) := e_{r_0^{-1}(n)}\). In particular, in the application of Theorem 4.11 (or more precisely, Corollary 4.13) in the proof of Theorem 4.14 above, this list is:
\[ e_1, C e_1, C^* e_1, \]
\[ C g_2, C g_3, \]
\[ C^* g_2, C^* g_3, C^* g_4, \]
\[ e_2, C g_4, C g_5, C g_6, C g_7, C g_8, C g_9, \]
\[ C^* g_5, C^* g_6, C^* g_7, C^* g_8, ..., C^* g_{15}, \]
\[ e_3, C g_{16}, C g_{17}, \ldots. \]

where \(g_2 = C e_1, g_3 = C^* e_1, g_4 = C g_2 = C^2 e_1, g_5 = C g_3 = C C^* e_1, \ldots\)

This is the same list which was constructed in the proof of [18, Theorem 20.8] for the linearly independent case. So the induction proof given above subsumes this linearly independent case.

Remark 4.16. Both Theorem 4.9 and Theorem 4.14 exhibit a lack of symmetry regarding the role of blocks \(\{A_n\}\) and \(\{B_n\}\). We note that it should be possible to obtain a more symmetric sparsification of \(A_n, B_n\) by modifying the functions \(r_0, r_1, r_2\) in Theorem 4.14.

In particular, set \(r_0(n) = 2^{n+1} - n - 2\), and define \(r_1(1) = 2\) and for all \(r_0(k - 1) < n \leq r_0(k)\) with \(k > 1\) (one can show \(k = \lceil \log_4(2n) \rceil - 1\)), we set
\[ r_1(n) := r_0(k) + 2(n - r_0(k - 1)) - 1, \]
\[ r_2(n) := r_1(n) + 1. \]

Note that \(r_0(n)\) satisfies the recurrence relation
\[ r_0(n+1) - r_0(n) = 2^{n+1} - 1 = 2(2^n - 1) + 1 = 2(r_0(n) - r_0(n - 1)) + 1. \]

Then it is not hard to check that these satisfy the hypotheses of Theorem 4.11.

Therefore, taking \(T_1 := C\), and \(T_2 := C^*\), then as in the proof of Theorem 4.14 in the basis \(\{f_n\}\) provided by Theorem 4.11 the operator \(C\) has column and row support lengths given by \(\langle r_1(n) \rangle, \langle r_2(n) \rangle\), respectively. Notice that
\[ r_2(r_0(n)) = r_0(n) + 2(r_0(n) - r_0(n - 1)) = r_0(n + 1). \]

Consequently, because \(r_i(r_0(n)) \leq r_0(n+1)\) for all \(i = 1, 2\) and \(n \in \mathbb{N}\), \(C\) is in block tridiagonal form with block sizes given by \(k_1 = 1 = r_0(1)\) and \(k_n = r_0(n) - r_0(n - 1) = 2^n - 1\) for \(n > 1\). In this case, the off-diagonal blocks, which have an aspect ratio barely exceeding 2 : 1, are more symmetrical. Indeed, the off-diagonal blocks are “upper (lower) triangular” in the sense that the \((i, j)\) entry of an off-diagonal
block is zero if \( i > 2j + 1 \) (respectively, \( j > 2i + 1 \)), i.e., the row and column support lengths increase by 2 instead of 1.

**Remark 4.17.** Reiterating our [18, Problem 20.11]: Given an operator \( C \), is there an orthonormal basis in which \( C \) is of the form as in Theorem 4.9, but where additionally each \( A'_{n} \) is a diagonal matrix? In particular, we wish to target the \( 5 \times 5 \) case as we did in [18].

5. Support density

In Section 4 we presented several universal block tridiagonal forms. For each of them we strove to find a basis in which the matrix of a given operator is sparse and had a particularly useful form. Now we will try to quantify their sparseness by considering the density of their support entries (meaning entries that need not be zero). This is a basis dependent measure that depends also on the index ordering of the basis having index order type \( \mathbb{N} \), i.e., order isomorphic to \( \mathbb{N} \).

Matrices can often carry data in a kind of organized way which we think of as special matrix forms. By a *matrix form*, we mean a subset of entry positions \((i,j)\) where matrices in the specified form must be zero; we refer to the positions in the complement of this set as *support entries*. For example, diagonal matrices are a matrix form corresponding to the subset \( \{(i,j) \mid i \neq j\} \). Note that any given matrix in a specified form may have zeros as some of its support entries.

Laurent and Toeplitz infinite matrices carry data that is spread out, whereas diagonal matrices or band matrices have their data relegated close to the diagonal, and these matrices are quite sparse. Similarly, any rank one projection operator can have a full matrix representation; but in some basis that rank one projection has all its support data relegated to its \((1,1)\) entry. This indicates the inherent basis dependent nature of these matrix forms. Our staircase and block tridiagonal forms also capture this idea by concentrating the support data, and our improvements at the end of the Section 4 sparsify these forms further still. In this section we attempt to measure sparsification as compared to upper triangular matrices, upper Hessenberg matrices, diagonal matrices, etc.

Given a finite \( N \times N \) matrix form, we define the number \( L_{N} \) as the number of support entries in it, and the support density by the average \( D_{N} = L_{N}/N^{2} \). For an infinite matrix form \( A \) we will consider the upper left \( N \times N \) corners, i.e., the compressions \( A_{N} := P_{N}AP_{N} \) with projections \( P_{N} \) onto \( \bigvee_{n=1}^{N} e_{n} \).

The support density, or simply called density, of \( A \) will mean

\[
D = \lim_{N \to \infty} D_{N} = \lim_{N \to \infty} \frac{L_{N}}{N^{2}},
\]

(5.1)

to measure the density of the support data, and our improvements at the end of the Section 4 sparsify these forms further still. In this section we attempt to measure sparsification as compared to upper triangular matrices, upper Hessenberg matrices, diagonal matrices, etc.

For several special classes of operators we briefly describe their special matrix forms and their densities. By the spectral theorem for normal compact operators, such an operator can be represented by a diagonal matrix, and the density of the diagonal matrix form is obviously zero. Moving away from the class of normal operators to the class of triangularizable operators, in the basis which triangularizes them all have density \( 1/2 \). Operators with a cyclic vector can be represented as an upper Hessenberg matrix. Indeed, for cyclic vector \( e \) apply the Gram–Schmidt process to \( \langle e, Te, T^{2}e, \ldots \rangle \). The upper Hessenberg matrix form is visually similar to an upper triangular matrix except that its support entries on the first subdiagonal. (In contrast, only triangularizable operators have bases relative to which their matrix representation is triangular.) Upper Hessenberg matrices have density \( 1/2 \), so all operators with cyclic vectors have a basis relative which their matrix representation has density \( 1/2 \). In the special case when the operator is selfadjoint the upper Hessenberg form reduces to a tridiagonal matrix, and the density of the tridiagonal matrix form (corresponding to \( \{|i,j| \mid |i - j| > 1\} \)) is clearly zero. Without a cyclic vector the construction used to obtain the last two matrix forms (Hessenberg and tridiagonal) leads to a direct sum of these forms, but our definition of density is not compatible with direct sums of infinite matrices.
With density defined above we can go back to universal forms of Section 4. We begin with the staircase form (4.1).

**Theorem 5.1.** The staircase form (4.1) has density $2/3$.

**Proof.** Recall that $(i, j)$ is a support entry of the staircase form (4.1) if and only if $i \leq 3j$ and $j \leq 3i$. Taking the $N \times N$ compression, the diagram in Figure 1 shows that the density $D_N = L_N/N^2$ clearly approaches $2/3$. Of course, the values of $L_N$ given in the diagram are not exact since they may not be integers, but they are at least asymptotically equivalent (they differ from the true values by at most $6N$), which is all that matters. □

![Figure 1. Asymptotic density of the staircase form (4.1).](image)

It was noticed in [18, Remark 20.3(i)] that, when $T$ has a cyclic vector (or more generally a vector $v$ for which the collection $p(T, T^*)v$ is dense), we obtain the matrix form where $(i, j)$ is a support entry if and only if $i \leq 2j$ and $j \leq 2i + 1$. For such matrices density is naturally smaller. Since the proof follows the same template we leave it to the reader and merely state the result.

**Theorem 5.2.** Every operator with cyclic vector has a basis relative to which its matrix is in staircase form with support rows and columns respectively $2, 4, 6, 8, \ldots$ and $3, 5, 7, \ldots$, and these matrices have density $1/2$.

In [18, Theorem 20.8] (Theorem 4.14 herein) we gave a modification of the construction of Theorem 4.1. It turns out that the sparsified matrices obtained that way have also density $1/2$ even when the operators from which they came do not have a cyclic vector.

**Theorem 5.3.** The sparsified staircase form obtained in the proof of Theorem 4.14 has density $1/2$. More generally, the staircase form of Corollary 4.13 has density $1/2$ whenever $\lim_{n \to \infty} n/r_0(n) = 0$.

**Proof.** Consider the staircase form of Corollary 4.13 and let $L_N$ denote the number of support entries in the upper left $N \times N$ corner with the convention that $L_0 := 0$. Then for $N \geq 0$ set $\Delta_N := L_{N+1} - L_N$ to be the number of support entries appearing in column $N + 1$ and row $N + 1$ (within the $(N + 1) \times (N + 1)$ block).
Of course, \( \Delta_0 = 1 \), and because \( T \) is a staircase form \( \Delta_N \leq \Delta_{N-1} + 2 \) for all \( N \geq 1 \). This is because support entries in subsequent rows/columns of a staircase form can only occur below and/or to the right of those in the previous row/column.

In fact, since the staircase form arises from Corollary 4.13, \( \Delta_N = \Delta_{N-1} + 2 \) if and only if \( N \in r_0(N) \), and otherwise \( \Delta_N = \Delta_{N-1} + 1 \). Indeed, if \( N = r_1(m) \), then the \((N + 1, m)\) position is not a support entry, and if \( N = r_2(m) \), then the \((m, N + 1)\) position is not a support entry, whereas if \( N = r_0(m) \), then both \((N + 1, m)\) and \((m, N + 1)\) are support entries. Therefore, the sequence \( \langle \Delta_N \rangle \) is the subsequence of positive integers which skips an integer at the indices \( r_0(N) \) for each \( N \).

Consequently, let \( N \in \mathbb{N} \) and suppose that \( k \) is the unique positive integer such that \( r_0(k) \leq N < r_0(k + 1) \). By filling in the missing integers (namely, \( r_0(j) + j \) for \( 1 \leq j \leq k \)) in the sequence \( \langle \Delta_N \rangle \) and taking the sum, we obtain

\[
L_N + \sum_{j=1}^{k} (r_0(j) + j) = \sum_{i=0}^{N-1} \Delta_i + \sum_{j=1}^{k} (r_0(j) + j) = \sum_{j=1}^{N+k} j = (N+k)(N+k+1)/2.
\]

Therefore,

\[
L_N = \frac{N^2 + 2kN + N}{2} - \sum_{j=1}^{k} r_0(j).
\]

Since \( 0 < r_0(j) \leq N \) for all \( 1 \leq j \leq k \), we find

\[
\frac{N^2 + kN + N}{2} \leq L_N < \frac{N^2 + 2kN + N}{2},
\]

and dividing through by \( N^2 \) yields

\[
\frac{1}{2} + \frac{k + 1}{2N} \leq D_N < \frac{1}{2} + \frac{2k + 1}{2N}.
\]

Finally, by hypothesis, as \( N \to \infty \) so also \( k \to \infty \), and \( k/N \leq k/r_0(k) \to 0 \). Hence \( D_N \to 1/2 \). \( \square \)

**Remark 5.4.** The previous theorem applies to the staircase form obtained in the proof of Theorem 4.14 but not to the block tridiagonal form specified in the statement of the theorem, even taking into account the sparsification obtained for the off-diagonal blocks. In fact, the density of the block tridiagonal form given in Theorem 4.14 can be shown to be \( 11/18 \). The discrepancy is due to the fact that block tridiagonal form does not take full advantage of the row support lengths dictated by the corresponding staircase form, as can be seen by examining the proof.

One could pose more general definitions of density for which the work above gives little information. Moreover, the density is highly basis dependent. We now show that even allowing permutations of certain bases, including those corresponding to our universal block tridiagonal form, can result in radically different densities, including density zero.

**Theorem 5.5.** Suppose that \( T \) is a matrix form and \( \langle e_n \rangle \) is an orthonormal basis of \( \mathcal{H} \). Suppose there is a subsequence \( \langle e_{m_k} \rangle_{k=1}^{\infty} \) such that the compression of \( T \) to \( \bigvee_{j=1}^{k} e_{m_j} \) has density \( \delta \). Then there is a permutation \( \pi \) of \( \mathbb{N} \) for which \( T \) has density \( \delta \) relative to the orthonormal basis \( \langle e_{\pi(n)} \rangle \).

**Proof.** Let \( \langle e_{n_k} \rangle \) be the complement subsequence of \( \langle e_{m_k} \rangle \). Let \( P_N \) denote the projection onto \( \bigvee_{j=1}^{N} e_{m_j} \). Let \( L_N \) denote the number of support entries of the compression of \( T \) to \( P_N \mathcal{H} \). Of course, by hypothesis \( L_N/N^2 \to \delta \) as \( N \to \infty \).
Now define a permutation \( \pi \) of \( \mathbb{N} \) by
\[
\pi(n) := \begin{cases}
  l_k, & n = 2^{k-1}, \\
  m_{n-k}, & 2^{k-1} < n < 2^k.
\end{cases}
\]

Let \( Q_N \) denote the projection onto \( \bigvee_{j=1}^N e_{\pi(j)} \). Notice that if \( 2^{k-1} \leq N < 2^k \) (i.e., \( k = \lfloor \log_2(N) \rfloor + 1 \)), then
\[
\{e_{\pi(1)}, \ldots, e_{\pi(N)}\} = \{e_{1}, \ldots, e_k\} \cup \{e_{m_1}, \ldots, e_{m_{N-k}}\}.
\]

Then let \( L_N' \) denote the number of support entries of the compression of \( T \) to \( Q_N \), then
\[
L_{N-k} \leq L_N' \leq L_{N-k} + 2kN.
\]

Notice that \( \lim_{N \to \infty} \frac{N-k}{N} = 1 \), and hence \( \lim_{N \to \infty} \frac{L_{N-k}}{N^2} = \delta \). Consequently, by this and the above display, \( L_N'/N^2 \to \delta \) as \( N \to \infty \), and so \( T \) has density \( \delta \) in the basis \( (e_{\pi(n)}) \).

**Corollary 5.6.** Suppose that \( T \) is a matrix form and \( (e_n) \) is an orthonormal basis of \( \mathcal{H} \) in which the support lengths of all the rows and columns of \( T \) are finite. Then there is a permutation \( \pi \) of \( \mathbb{N} \) for which \( T \) has density 0 relative to the basis \( (e_{\pi(n)}) \).

**Proof.** Since all the rows and columns of \( T \) have finite support lengths, there is a subsequence \( (e_{m_k}) \) for which the compression of \( T \) to \( \bigvee_{k=1}^\infty e_{m_k} \) is diagonal, and therefore has density 0 relative to this basis for the subspace. By Theorem 5.5, there is a permutation \( \pi \) of \( \mathbb{N} \) for which \( T \) has density 0 relative the basis \( (e_{\pi(n)}) \).

**Corollary 5.7.** For every operator \( T \in \mathcal{B}(\mathcal{H}) \), there is some orthonormal basis \( (e_n) \) in which \( T \) has density 0.

**Proof.** By Corollary 5.6 we may take the basis to be a permutation of the basis arising from any staircase form for \( T \).

6. **Constraints on the Pearcy–Topping problem from s-numbers and \( \mathcal{B}(\mathcal{H}) \)-ideals**

Much of this paper is focused on some internal matrix structure constraints on the search for compacts \( C, Z \) when solving \( T = CZ - ZC \). In this section, we provide some elementary s-number quantitative constraints (that more generally apply to finite sums of bi-products of operators) and some consequent \( \mathcal{B}(\mathcal{H}) \)-ideal algebraic qualitative constraints with some examples and some problems. For these we need the characterization: compact operators are those whose \( s \)-number approach provides necessary quantitative \( s \)-number sequence tends to 0. Unlike the past sections concentrating on the innards of certain matrix representations, some of the conditions presented in this section are independent of bases representations of the operators \( C, Z \). We also provide an s-number and ideal analysis of Anderson’s original contruction for the rank one projection \( P = CZ - ZC \) along with an analysis of the relationship between his \( C \) (where \( C = C_1 + C_{-1} \)) and its two block upper and lower diagonals \( C_1, C_{-1} \) and their principal ideals \( (C), (C_1), \) and \( (C_{-1}) \), and similarly for \( Z \). This analysis originated in [18]. A direct s-number approach provides necessary quantitative s-number conditions for \( T = CZ - ZC \) (see Proposition 6.1).

We will not assume that the reader is familiar with the theory of \( \mathcal{B}(\mathcal{H}) \)-ideals or s-numbers. The s-numbers of a compact operator \( T \) are defined as \( s_n(T) := \inf\{\|T - F\| \mid \text{rank } F < n \} \) for \( n \in \mathbb{N} \), and it is well-known that the sequence \( s(T) := (s_n(T))_{n=1}^\infty \) lists the eigenvalues of \( |T| \) with multiplicity in nonincreasing order (omitting the eigenvalue 0 if \( T \) has infinite rank). Unless otherwise indicated, all operations on sequences described in this section are to be performed pointwise, e.g., \( \sqrt{s(T)} := (\sqrt{s_n(T)}) \).

Given sequences \( \lambda, \mu \), we let \( \lambda \lor \mu, \lambda \land \mu \) denote their pointwise maximum and minimum, respectively.

\(^3\)For the proof, it suffices that the permutation \( \pi \) satisfies \( k/\pi^{-1}(l_k) \to 0 \) and both \( \pi^{-1}(l_k) \) and \( \pi^{-1}(m_k) \) are strictly increasing.
For this section we will recall the Calkin–Schatten characterization for $\mathcal{B}(\mathcal{H})$-ideals in terms of characteristic sets (see for instance [6, Section 2.7 end]). To any proper $\mathcal{B}(\mathcal{H})$-ideal $\mathcal{J}$ one can associate its characteristic set

$$\Sigma(\mathcal{J}) := \{s(T) \mid T \in \mathcal{J}\}.$$ 

Then $\Sigma(\mathcal{J})$ is a subcone of the cone of nonnegative nonincreasing sequences converging to zero, and moreover this subcone is hereditary (under the pointwise order) and closed under the $k$-ampliation operator $D_k(\langle a_n \rangle) := \langle a_{\lfloor n/k \rfloor} \rangle$. Likewise, given any such subcone $\Sigma$, one can define the associated ideal

$$\mathcal{J}_\Sigma := \{T \in \mathcal{K}(\mathcal{H}) \mid s(T) \in \Sigma\}.$$ 

The map $\mathcal{J} \mapsto \Sigma(\mathcal{J})$ is an order (and semiring) isomorphism from the (subset) lattice of proper operator ideals into the (subset) lattice of hereditary subcones closed under 2-ampliation, and its inverse is $\Sigma \mapsto \mathcal{J}_\Sigma$. For details consult [20, p. 26, Theorem 12]

The Calkin–Schatten characterization often significantly simplifies analysis in the theory of operator ideals and allows for ideal constructions by defining the corresponding construction on characteristics sets; this is a key feature often exploited in [6]. As an example, any operator ideal $\mathcal{J}$ has a natural well-defined square root $\mathcal{J}^{1/2}$ whose characteristic set is given by $\Sigma(\mathcal{J}^{1/2}) := \{\sqrt{s(T)} \mid s(T) \in \Sigma(\mathcal{J})\}$, which is a hereditary subcone closed under ampliations.

Recall two basic $s$-number inequalities [11, Corollary 2.2]:

$$s_{n+m-1}(AB) \leq s_n(A)s_m(B) \quad \text{and} \quad s_{n+m-1}(A + B) \leq s_n(A) + s_m(B),$$ 

which by straightforward induction arguments establish:

$$s_{m_1 + \cdots + m_k - (k-1)} \left( \prod_{i=1}^{k} A_i \right) \leq \sum_{i=1}^{k} s_{m_i}(A_i) \quad \text{and} \quad s_{m_1 + \cdots + m_k - (k-1)} \left( \sum_{i=1}^{k} A_i \right) \leq \sum_{i=1}^{k} s_{m_i}(A_i).$$

For $m \in \mathbb{N}$, we may divide $m$ by $k$ to get a quotient $q = \lceil m/k \rceil$ with nonpositive remainder $-r$ with $0 \leq r < k$, so that $m = qk - r \geq k - (k-1)$. Then from the above display, we obtain:

$$s_m(A_1 \cdots A_k) = s_{qk-r}(A_1 \cdots A_k) \leq s_{qk-(k-1)}(A_1 \cdots A_k) \leq s_q(A_1) \cdots s_q(A_k),$$ 

and similarly for $s_n(A_1 + \cdots + A_k)$. Therefore since $q = \lceil m/k \rceil$ and $D_k(\langle a_m \rangle) = \langle a_{\lfloor m/k \rfloor} \rangle$,

$$s(A_1 \cdots A_k) \leq D_k(s(A_1) \cdots s(A_k)) \quad \text{and} \quad s(A_1 + \cdots + A_k) \leq D_k(s(A_1) + \cdots + s(A_k)). \quad (6.1)$$

A quantitative $s$-number sequence analysis for $T = CZ - ZC$.

Clearly, $T \in (CZ) + (ZC) \subseteq (C)(Z)$ is a qualitative constraint, but an explicit quantitative constraint is:

**Proposition 6.1.** If $T = \sum_{i=1}^{k} C_i Z_i$, with each $C_i, Z_i$ compact, then

$$s(T) \leq \sum_{i=1}^{k} D_{2k}(s(C_i)s(Z_i)).$$

In particular, if $T = CZ - ZC$ for $C, Z$ both compact, then $s(T) \leq 2D_4(s(C)s(Z)).$

**Proof.** By (6.1), and noting that the ampliation operations are semiring endomorphisms and $D_{2k} = D_k \circ D_{2k}$,

$$s(T) = s \left( \sum_{i=1}^{k} C_i Z_i \right) \leq D_k \left( \sum_{i=1}^{k} s(C_i Z_i) \right) \leq D_k \left( \sum_{i=1}^{k} D_2(s(C_i)s(Z_i)) \right) = D_{2k} \left( \sum_{i=1}^{k} s(C_i)s(Z_i) \right).$$

For $T = CZ - ZC$, since $s(-Z) = s(Z)$, the above yields $s(T) \leq 2D_4(s(C)s(Z)).$\hfill $\square$

---

4Since $\Sigma(\mathcal{J})$ is hereditary, it is closed under 2-ampliation if and only if it is closed under $k$-ampliation for any $k \geq 2$. 

---
Remark 6.2. Since Proposition 6.1 applies equally well to \( T = CZ + ZC \), and even more generally to finite sums of bi-products, this proposition does not distinguish commutators on its own.

In any case, recalling historical facts on sums of commutators: Every bounded operator is the sum of two commutators of bounded operators (Brown–Pearcy and Brown–Halmos–Pearcy, see [19, p. 248, paragraph below Problem 1, and citations]); Every compact operator is the finite sum of commutators of bounded operators (Brown–Pearcy and Brown–Halmos–Pearcy, see [19, p. 248, paragraph below Problem 1, and citations]); Every compact operator is the finite sum of commutators of bounded operators (Brown–Pearcy and Brown–Halmos–Pearcy, see [19, p. 248, paragraph below Problem 1, and citations]); every compact operator is the sum of two commutators of compact operators, i.e., \( K(H) = [K(H), K(H)] \) [19 Theorem 1]; and from this in [6, Corollary 6.5] it follows that every compact operator is the sum of four commutators, i.e., \( K(H) = [K(H), K(H)]_4 \), although [6, Corollary 6.5] applies more generally to the commutator space \( [I, J] = [I, J]_4 \) for arbitrary operator ideals \( I, J \). For \( K(H) \) specifically, in fact even more is known: every compact operator is the sum of two commutators of compact operators, which is due to Fan and Fong in [10].

So for \( T \in K(H) \), one has \( T = \sum_{i=1}^{2} [C_i, Z_i] \). For these operators applying Proposition 6.1 one obtains

\[
s(T) \leq 2D_s(s(C_1)s(Z_1) + s(C_2)s(Z_2)).
\]

(6.2)

While we don’t know even for all strictly positive operators \( T \) that a representation as a single commutator exists (although we provided a class of such strictly positive compact operators in Section 2), we note that if \( T = CZ - ZC \), then Proposition 6.1 affords the alternate constraint

\[
s(T) \leq 2D_4(s(C)s(Z)).
\]

(6.3)

A qualitative analysis of \( B(H) \)-ideal constraints and open questions. The previous quantitative sharpening of the constraints has no bearing on the following ideal analysis, but the qualitative implications of the \( s \)-number inequalities do not.

A purely algebraic approach tells us: \( T = CZ - ZC \Rightarrow T \in (C)(Z) \); indeed, ideal product is a commutative operation, and also \( (T) \subseteq (C)(Z) = (\text{diag}(s(C)s(Z))) \) (see [6] Section 2.8 and Theorem 6.3).

Thus we have the following interesting question for compact \( C, Z \):

Question 2. When is it possible to choose \( C, Z \) such that \( T = CZ - ZC \) and \( (T) = (C)(Z) \)?

For the remainder of this subsection, we analyze several different natural scenarios. Note that if \( T = [C, Z] \), then \( (T) \subseteq (C)(Z) \). In general, any of these ideal inclusions may be proper. We note that there are known instances in which equality (even \( (T) = (C) \)) can be achieved, but also established instances for which the containment is forced to be proper (even \( (T) \nsubseteq (C)(Z) \)).

Example 6.3. Given any sequence \( \langle d_n \rangle \), the operator \( T = \text{diag} \langle d_1, -d_1, d_2, -d_2, \ldots \rangle \) can be written as the commutator

\[
T = [C, Z] := \left[ \bigoplus \sqrt{d_n} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \bigoplus \sqrt{d_n} \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \right].
\]

(6.3)

(This is an example of broader phenomena, selfcommutators \( [A, A^*] = A^*A - AA^* \) of a compact operator \( A \) which have been characterized by P. Fan and C. Fong [19 Theorem 1].) Of course, when \( \langle d_n \rangle \) is nonnegative nonincreasing and converging to zero, then \( T \) is compact, and moreover, one can check that \( s(T) = D_2(\langle d_n \rangle) \) and \( s(C) = s(Z) = (\sqrt{d_n}) \). Consequently, by appeal to characteristic sets, we easily obtain \( (T) = (C)(Z) \).

In fact, certain choices of \( \langle d_n \rangle \), for example \( d_n = 2^{-n} \), even yield \( (T) = (C) = (Z) \). Indeed, in this case, notice that \( \sqrt{d_n} = 2^{-n/2} \leq 2^{-[n/2]} = 2^{1/2} \cdot 2^{-[n/2]} \), and hence \( s(C) = s(Z) \leq \sqrt{2} 2^{-[n/2]} = \sqrt{2}s(T) \). Thus, \( C, Z \in (T) \), hence equality of the corresponding ideals. Of course, in general, \( (C)(Z) \subseteq (C) \cap (Z) \) is a proper inclusion (e.g., \( d_n = 1/n \), because \( s(C) \neq O(D_k(s(T))) \) for any \( k \)), so we would rarely expect these three principal ideals to be equal, or even merely \( (C)(Z) = (C) \cap (Z) \).
In the previous example, whenever the operator $T$ is trace-class (i.e., $\langle d_n \rangle$ absolutely summable), its trace is zero. This may lead one to wonder about trace-class $T$ with nonzero trace. We now concern ourselves primarily with this situation.

**Example 6.4.** Suppose that $T = [C, Z]$ is trace-class with $C, Z \in B(H)$. If either of $CZ, ZC$ are trace-class, then both would be (since $T$ is), and hence [15] Lemma 2.1 guarantees $\text{tr}(CZ) = \text{tr}(ZC)$, so $\text{tr}(T) = \text{tr}([C, Z]) = 0$. Therefore, if $\text{tr}(T) \neq 0$, then we must have $CZ, ZC \notin C_1$. Consequently, $T$ trace-class with $\text{tr}(T) \neq 0$ implies $(T) \not\subseteq (CZ) \subseteq (C)(Z)$.

**Remark 6.5.** A sufficient condition for $CZ, ZC \in C_1$ (and hence $\text{tr}(T) = \text{tr}([C, Z]) = 0$) is $(C) \cap (Z) \subseteq C_1$. One may wonder whether the inclusion $(C) \cap (Z) \subseteq C_1$ is even possible unless either $C$ or $Z$ is itself trace-class. We remark that this is indeed possible even if $C, Z \notin C_1$.

First, it is elementary in the theory of operator ideals using characteristic sets to prove that $(C) \cap (Z) = (\text{diag}(s(C) \wedge s(Z)))$. So, it suffices to exhibit nonsummable sequences $c, z \in c_0^N (= s(C), s(Z))$ for which $c \wedge z$ is summable.

We construct $c, z$ as follows. For $k \in \mathbb{N}$, set $n_k := k(k + 1)/2$, and set $s_k := \sum_{j=1}^k 2^n$ and for convenience set $s_0 := 0$. Then define,

$$c_m := 2^{-n_{k+1}} \text{ for } s_{2(k-1)} < m \leq s_{2k}, k \in \mathbb{N}$$

$$z_m := \begin{cases} 1 & 1 \leq m \leq s_1 \\ 2^{-n_{k+1}} & s_{2k-1} < m \leq s_{2k+1}, k \in \mathbb{N} \end{cases}$$

Notice that, by construction, 

$$(c \wedge z)_m = 2^{-n_{k+1}} \text{ for } s_{k-1} < m \leq s_k, k \in \mathbb{N}.$$ 

Therefore,

$$\sum_{m=1}^\infty (c \wedge z)_m = \sum_{k=1}^\infty \sum_{m=s_{k-1}+1}^{s_k} (c \wedge z)_m = \sum_{k=1}^\infty \sum_{m=s_{k-1}+1}^{s_k} 2^{-n_{k+1}} = \sum_{k=1}^\infty 2^n 2^{-n_{k+1}} = \sum_{k=1}^\infty 2^{-k} = 1.$$ 

On the other hand,

$$\sum_{m=1}^\infty c_m = \sum_{k=1}^\infty \sum_{m=s_{k-1}+1}^{s_{2k}} 2^{-n_{k+1}} = \sum_{k=1}^\infty (2^{n_{2k-1}} + 2^{n_{2k}}) 2^{-n_{k+1}} \geq \sum_{k=1}^\infty 2 = \infty,$$

and

$$\sum_{m=1}^\infty z_m \geq \sum_{k=1}^\infty \sum_{m=s_{k-1}+1}^{s_{2k+1}} 2^{-n_{k+1}} = \sum_{k=1}^\infty (2^{n_{2k}} + 2^{n_{2k+1}}) 2^{-n_{k+1}} \geq \sum_{k=1}^\infty 2 = \infty.$$ 

**Example 6.6.** The earlier Example 6.4 implies that if $C \in C_p, Z \in C_q$ (Schatten ideals) and $T = [C, Z]$ has nonzero trace, then $1/p + 1/q < 1$ (or equivalently, if $1/p + 1/q \geq 1$, then $T = [C, Z]$ has zero trace). Rather interestingly, Brown proved [4] that when the commutator $T = [C, Z]$ is finite rank, then even more can be said, namely, if $1/p + 1/q \geq 1/2$ (hence $CZ, ZC \in C_2$), then $\text{tr}(T) = 0$.

In case $P$ is a rank-one projection and $P = [C, Z]$ then at least one of $C, Z \notin C_p$ for all $p \leq 4$. In Anderson’s construction with $t = 1/2$, [11] Theorem 3] establishes $C, Z \in \bigcap_{p>4} C_p$, so in some sense it appears that Anderson’s construction may be optimal. In Lemma 6.7 below, we compute explicitly the principal ideal $(C) = (Z)$ arising from Anderson’s construction with $t = 1/2$. It would be interesting to know if Anderson’s construction really is minimal (see Question 5).

The above examples lead us to pose the following questions:

**Question 3.** If $T = [C, Z]$ is positive and not trace-class, is the inclusion $(T) \not\subseteq (C)(Z)$ always proper?
Question 4. If \( T = [C, Z] \) is trace-class and \( C, Z, ZC \in C_2 \), do these conditions imply \( \text{tr}(T) = 0 \)?

Question 3 is motivated by Examples 6.3 and 6.4. In particular, Example 6.3 shows that in order to guarantee the inclusion is proper in the case when \( T = [C, Z] \) is not trace-class, one needs some additional hypotheses. Given that all the operators \( T \) from Example 6.4 in some sense have “trace zero” (because the positive and negative parts are unitarily equivalent), Example 6.4 suggests the natural test case is when this is asymmetric, i.e., \( T \) is positive and not trace-class.

Question 4 is motivated from Example 6.6 and [4]. In the proof in [4], the Schatten class assumption of the individual operators \( C, Z \) was only used to bound Hilbert–Schmidt norms associated to submatrices of the products \( CZ, ZC \), suggesting that maybe there is another approach which replaces this assumption only with \( CZ, ZC \in C_2 \). Likewise, the finite rank assumption on \( T \) was already partially loosened in [4] to \( T \) trace-class with \([C, C^*]\) finite rank.

Before we compute explicitly the principal ideals \((C), (Z)\) arising from Anderson’s construction for the rank-one projection \( P = [C, Z] \) for a rank-one projection \( P \), we first recall for the reader some basic facts about nonnegative nonincreasing sequences converging to zero which will be used below. For readers intimately familiar with the theory of operator ideals, these facts will be well-known; for others, they should be comprehensible with a bit of thought.

Suppose \( \lambda = (\lambda_n), \mu = (\mu_n) \) are positive sequences converging to zero and suppose \( \lambda \leq \mu \). Then if \( \lambda^\uparrow, \mu^\uparrow \) denote their nonincreasing rearrangements, \( \lambda^\uparrow \leq \mu^\uparrow \) (when a sequence has infinite support, any zeros in the sequence are omitted in its nonincreasing rearrangement). Moreover, the collection \( c^\uparrow \) of nonnegative nonincreasing sequences converging to zero has an internal direct sum (cf. [6] §2.4) operation given by \( \lambda, \mu \in c^\uparrow \mapsto (\lambda \oplus \mu)^\uparrow \), where \( (\lambda \oplus \mu)_{2n} := \lambda_n \) and \( (\lambda \oplus \mu)_{2n+1} := \mu_n \). We note that \( \lambda \vee \mu \leq (\lambda \oplus \mu)^\uparrow \leq D_2(\lambda \vee \mu) \).

In what follows, we will also be concerned with finite sequences. Suppose that for each \( n \in \mathbb{N} \), \( a_n \) denotes some finite sequence. Then \( \bigoplus_n a_n \) is the concatenation of these sequences. If the finite sequences \( a_n \) are nonnegative and, for any \( \epsilon > 0 \), there are only finitely many \( n \) such that any term in \( a_n \) exceeds \( \epsilon \), then \( \bigoplus_n a_n \) is nonnegative and converges to zero. In this case, we note that the monotonization and \( k \)-ampliation operators \( (\cdot)^\uparrow \) and \( D_k(\cdot) \) are well-behaved with respect to the direct sum. More precisely,

\[
D_k \left( \left( \bigoplus_n a_n \right)^\uparrow \right) = \left( \bigoplus_n D_k(a_n) \right)^\uparrow.
\]

Of course, \( D_k \) applied to a sequence of length \( m \) produces a sequence of length \( km \).

**Lemma 6.7.** For Anderson’s commutator construction for the rank-one projection \( P = CZ - ZC \), \( (C) = (Z) = (\text{diag}(1/\sqrt{n})) \).

**Proof.** In the notation of Theorem 6.8 below, \( C_{\pm 1}, Z_{\pm 1} \) denote the upper and lower block diagonal operators of \( C, Z \), respectively. Moreover, by Theorem 6.3, we have \( (C) = (C_1) + (C_{-1}) \) because \( C_0 = 0 \), and similarly for \( Z \). A simple computation shows that \( C_{\pm 1}^* C_{\pm 1} \) and \( Z_{\pm 1}^* Z_{\pm 1} \) are diagonal, and hence these diagonal entries are precisely the singular values \( s(C_{\pm 1}^* C_{\pm 1}) = s(|C_{\pm 1}|^2) = s(C_{\pm 1})^2 \), and similarly for \( s(Z_{\pm 1}) \).

For each \( n \in \mathbb{N} \), consider the finite sequence \( c_n := (n - k + 1)/n^2 \) \( k = 1 \). Since none of the terms in \( c_n \) exceed \( 1/n \), it is clear that \( c := \bigoplus_n c_n \) is nonnegative and converges to zero. Examination of the definitions of \( C_{\pm 1}, Z_{\pm 1} \) (see [2] p. 7) or the original [11] pp. 129–130 after setting his parameter \( t = 1/2 \) shows that the nonzero diagonal entries of each of \( C_{\pm 1}^* C_{\pm 1}, Z_{\pm 1}^* Z_{\pm 1} \) are precisely given by the sequence \( c \) (up to permutation). Therefore, \( s(Z_{\pm 1})^2 = s(C_{\pm 1})^2 = c^\uparrow \), hence \( (Z_{\pm 1})^2 = (C_{\pm 1})^2 = (\text{diag} c^\uparrow) \).

And thus \( (Z) = (C) = (C_{-1}) + (C_1) = (C_1) = (\text{diag} \sqrt{c^\uparrow}) \).
Now, note that $2D_2(c_n)$ is a finite sequence of length $2n$, and we may let $d_n, d'_n$ denote the sequences of length $n$ so that $2D_2(c_n)$ is the concatenation $d_n \oplus d'_n$ of $d_n$ followed by $d'_n$. We remark that $d_n \geq d'_n$ (since $2D_2(c_n)$ is nonincreasing). We let $d := \bigoplus_n d_n$, $d' := \bigoplus_n d'_n$; clearly, $d \geq d'$ and so $d^\ast \geq d'^\ast$. Moreover, notice that $d_n \geq a_n := (1/n)_{k=1}^n$. Therefore, $d \geq a := \bigoplus_n a_n$, and so $d^\ast \geq a^\ast = a$.

Putting all this together yields

$$2D_2(c^\ast) = \left( \bigoplus 2D_2(c_n) \right)^\ast = \left( \bigoplus (d_n \oplus d'_n) \right)^\ast = (d \oplus d')^\ast = (d^\ast \oplus d'^\ast)^\ast \geq d^\ast \lor d'^\ast \geq d^\ast \geq a^\ast = a.$$ 

In addition, we remark that $c_n \leq a_n$, and hence $c \leq a$, so $c^\ast \leq a^\ast = a$. These inequalities (namely, $c^\ast \leq a$ and $a \leq 2D_2(c^\ast)$) guarantee that the characteristic sets generated by $c^\ast$ and $a$ coincide, or equivalently, the principal ideals $(\text{diag } c^\ast)$ and $(\text{diag } a)$ are equal.

Consider now the sequence $(1/\sqrt{n}) \in c^\ast$. We show that $(1/\sqrt{n})$ and $a$ generate the same characteristic set. Given $n \in \mathbb{N}$, there is some $k \in \mathbb{N}$ for which $k(k+1)/2 \leq n < (k+1)(k+2)/2$, and hence the $n$th term of $a$ lies in the finite constant sequence $a_k$, and therefore takes the value $1/k$. Dividing the inequalities through by $k^2$, we see that

$$\frac{1}{2} \geq \frac{1}{k^2} \leq \frac{n}{k^2} < \left( \frac{1}{2} + \frac{1}{2k} \right) \left( 1 + \frac{2}{k} \right) \leq 3.$$ 

Consequently, $a \leq \sqrt{3}(1/\sqrt{n})$ and $(1/\sqrt{n}) \leq \sqrt{2}a$, and therefore these sequences generate the same characteristic set.

Finally, we conclude

$$(C)^2 = (\text{diag } c^\ast) = (\text{diag } a) = (\text{diag } (1/\sqrt{n})), \quad (Z) = (C) = (\text{diag } \sqrt{c^\ast}) = (\text{diag } (1/\sqrt{n})).$$

Lemma 6.7 along with Example 6.6 lead us to pose the following question regarding minimality of the ideals $(C), (Z)$ arising from a commutator representation $P = [C, Z]$ of a rank-one projection.

**Question 5.** Let $P$ be a rank-one projection. Do there exist $C, Z$ with $P = [C, Z]$ for which both $(C), (Z) \subseteq (\text{diag } (1/\sqrt{n}))$ and at least one inclusion is proper?

In the previous question, we explicitly ask for both principal ideal inclusions. The reason for this is that it is possible to obtain an asymmetric solution were one of the principal ideal inclusions is proper, whereas the other doesn’t hold. Indeed, from Anderson [11, Theorem 3], it is clear that for any $1/2 < t < 1$, one may obtain operators $C_t, Z_t$ for which $P = [C_t, Z_t]$ and $(C_t) \subseteq C_{2/t} \subseteq (\text{diag } (1/\sqrt{n}))$. However, in this case one cannot have $Z_t \subseteq (\text{diag } (1/\sqrt{n})) \subseteq \bigcap_{p>4} C_p$ because it would violate the result of Brown [4] mentioned in Example 6.6.

This problem comes from similar ideas studied in [6, Section 7] and the open question there for ideals $I$: If $[I, \mathcal{B}(\mathcal{H})]_1$ (or $[I^{1/2}, I^{1/2}]_1$) contains $P$ (or some finite rank nonzero trace operator), must $\text{diag } (1/\sqrt{n}) \in I$? Also worthy of note is Peary–Topping’s [14, Problem 2] on Schatten-$p$ commutators: Is $[C_2, C_p]_1 = C_p$ for $p > 1$, to which Brown provided the negative solution [4] when $p \leq 2$. The case $p > 2$ remains open.

**Matrix splitting and $\mathcal{B}(\mathcal{H})$-ideals.** For this subsection we mention two special case theorems whose much more general versions are in [16, Theorems 4 and 8] auxiliary to this work, and Theorem 6.8 in particular was used in the proof of Lemma 6.7 above.

Our universal block tridiagonal matrices mentioned throughout this paper can be described by means of a sequence of orthogonal projections summing to the identity $(P_n)_{n \in \mathbb{N}}$ via its block diagonals $T_0, T_{\pm 1}, \ldots$.
that is, \( T = T_{-1} + T_0 + T_1 \) where \( T_0 := \sum_{n=1}^{\infty} P_n T P_n, T_1 := \sum_{n=1}^{\infty} P_n T P_{n+1} \) and \( T_{-1} := \sum_{n=1}^{\infty} P_{n+1} T P_n \).

The exponential dimensions of our projections are prescribed in Theorems 4.1, 4.3 and 4.4 but the next theorem for block tridiagonal matrices also extends naturally to block banded matrices of arbitrary bandwidth and block sizes.

**Theorem 6.8** ([16, Theorem 4]). Let \( T \in K(H) \) be represented in any of its universal block tridiagonal forms with \( T_0, T_{\pm 1} \) representing its central, upper and lower block diagonal matrices, so \( T = T_{-1} + T_0 + T_1 \). Then each \( T_0, T_{\pm 1} \in (T) \) and the following equality of ideals holds:

\[
(T) = (T_{-1}) + (T_0) + (T_1).
\]

An elementary consequence is: Let \( CZ - ZC = D \) with \( C \) compact and \( Z \) bounded.

(i) Then \( D \in (C) \), i.e., \( (D) \subseteq (C) \), or equivalently, \( s(D) = O(D_k s(C)) \) for some \( k \)-ampliation \( D_k \).

(ii) In any basis where \( C \) is block tridiagonal, i.e., \( C = C_{-1} + C_0 + C_1 \), one has \( C_{-1}, C_0, C_1 \in (C) \).

And so each have \( s \)-number sequences satisfying the above \( O \)-condition in (i) for some \( k \).

(iii) If we simultaneously block tridiagonalize \( C, D \) or all three operators (which can be achieved by Theorem 4.1), the same can be said for each of them as in (ii).

Regarding the Pearcy–Topping problem equivalences mentioned earlier (Theorem 4.6), we have

**Corollary 6.9.** For \( T = CZ - ZC \) with all three operators compact, when they are simultaneously put in universal block tridiagonal form relative to any basis, one has the constraints: \( (T) \subseteq (C)(Z) \) along with \( (T) = (T_{-1}) + (T_0) + (T_1) \). And the latter equality holds likewise for \( C \) and \( Z \).

Another mentioned equivalence (see Theorem 4.6) where diagonal \( D = CZ - ZC \), when \( C, Z \) are simultaneously put in universal block tridiagonal form relative to the basis diagonalizing \( D \), then \( D \) is a block 5-diagonal matrix and one has the constraints \( (D) \subseteq (C)(Z) \) where \( (D) = (D_{-2}) + (D_{-1}) + (D_0) + (D_1) + (D_2), (C) = (C_{-1}) + (C_0) + (C_1) \) and \( (T) = (Z_{-1}) + (Z_0) + (Z_1) \).

And for general interest, though not used in this paper, [16, Theorem 9] handles the more general case where \( T \) is a full block matrix, that is, not finite block banded, but the conclusion is not as strong as Theorem 6.8. This slightly weaker conclusion obtained in [16] is, in general, optimal and involves the notion of the arithmetic mean closure of the relevant ideal.

In the spirit of Sections 2 and 3 (esp., (2.1) and Theorem 3.1 regarding matrix finite full blocks), we close this matrix splitting subsection with a \( B(H) \)-ideal consequence of boundedness of \( k_n(||A_n|| + ||B_n||) \) for block tridiagonal \( T \) with zero central blocks.

**Proposition 6.10.** Suppose \( T \in B(H) \) be in block tridiagonal form with zero central blocks, i.e.,

\[
T = \begin{pmatrix}
0 & A_1 & 0 & \cdots \\
B_1 & 0 & A_2 & \ddots \\
0 & B_2 & 0 & \ddots \\
\vdots & \ddots & \ddots & \ddots \\
\end{pmatrix}, \quad \text{with central blocks of size } k_n \times k_n,
\]

and let \( r_n := \sum_{j=1}^{n} k_j \) denote the partial sums, and let \( r_0 := 0 \). If the sequences of norms \( ||A_n||, ||B_n|| \) are nonincreasing and the sequence \( \langle r_n(||A_n|| + ||B_n||) \rangle \) is bounded, then \( (T) \subseteq (\text{diag}(1/n)) \).

Moreover, if \( k_n = \Theta(\rho^n) \) for some \( \rho > 1 \), then the boundedness condition can be replaced by the boundedness of \( \langle k_n(||A_n|| + ||B_n||) \rangle \).

**Proof.** We begin by noting that since \( r_n \to \infty \), the boundedness of \( \langle r_n(||A_n|| + ||B_n||) \rangle \) guarantees that \( ||A_n||, ||B_n|| \to 0 \), and hence \( T \) is compact.

As in the proof of Lemma 6.7 we apply Theorem 6.8 to conclude that \( (T) = (T_1) + (T_{-1}) \), so it suffices to prove \( (T_{\pm 1}) \subseteq (\text{diag}(1/n)) \). By symmetry considerations, it suffices to establish this only for \( T_1 \). We
note, as in the proof of Lemma 6.7, that \( s(T_1) = (\bigoplus s(A_n))^\star \), and also \( s(A_n) \leq (\|A_n\|)^{k_n} \) (the latter is a constant sequence). Therefore,
\[
s(T_1) = \left( \bigoplus_{n} s(A_n) \right)^\star \leq a := \bigoplus_{n} \left( (\|A_n\|)^{k_n} \right).
\]

The sequence \( a = a^\star \) is already monotone since the sequences \( (\|A_n\|), (\|B_n\|) \) are nonincreasing.

Then let \( M > 0 \) denote a bound for \( \langle r_n(\|A_n\| + \|B_n\|) \rangle \). Then, for any \( m \in \mathbb{N} \), there is some \( n \in \mathbb{N} \) with \( r_{n-1} < m \leq r_n \), and by the definition of \( a \), \( a_m = \|A_n\| \). Thus
\[
a_m = \|A_n\| \leq \frac{M}{r_n} \leq \frac{M}{m}.
\]

Hence \( a \) (and also \( s(T_1) \)) lies in any characteristic set containing \( \langle 1/n \rangle \), so \( (T_1) \subseteq (\text{diag}(1/n)) \).

If, in addition, \( k_n = \Theta(\rho^n) \) for some \( \rho > 1 \), we note that \( k_n = \Theta(r_n) \) and hence \( \langle r_n(\|A_n\| + \|B_n\|) \rangle \) is bounded if and only if \( \langle k_n(\|A_n\| + \|B_n\|) \rangle \) is bounded. \( \square \)

References

[1] Joel Anderson. Commutators of compact operators. J. Reine Angew. Math., 291:128–132, 1977.
[2] Daniel Beltiță, Sasmita Patnaik, and Gary Weiss. \( B(H) \)-commutators: a historical survey II and recent advances on commutators of compact operators. In The varied landscape of operator theory, volume 17 of Theta Ser. Adv. Math., pages 57–75. Theta, Bucharest, 2014.
[3] Arlen Brown and Carl Pearcy. Structure of commutators of operators. Ann. of Math. (2), 82:112–127, 1965.
[4] L. G. Brown. Traces of commutators of Schatten-von Neumann class operators. J. Reine Angew. Math., 451:171–174, 1994.
[5] Kenneth R. Davidson, Laurent Marcoux, and Heydar Radjavi. A strictly positive commutator of compacts. preprint.
[6] Ken Dykema, Tadeusz Figiel, Gary Weiss, and Mariusz Wodzicki. Commutator structure of operator ideals. Adv. Math., 155(1):1–79, 2004.
[7] Ken Dykema and Amudhan Krishnaswamy-Usha. Nilpotent elements of operator ideals as single commutators. Proc. Am. Math. Soc., 146(7):3031–3037, 2018.
[8] Ky Fan. Maximum properties and inequalities for the eigenvalues of completely continuous operators. Proc. Nat. Acad. Sci., U. S. A., 37:760–766, 1951.
[9] Peng Fan and Che-Kao Fong. Which operators are the self-commutators of compact operators? Proc. Amer. Math. Soc., 80:58–60, 1980.
[10] Peng Fan and Che-Kao Fong. Operators similar to zero diagonal operators. Proc. R. Ir. Acad., Sect. A, 87(2):147–153, 1987.
[11] Israel C. Gohberg and Mark Grigor’evich Kreĭn. Introduction to the theory of linear nonselfadjoint operators, volume 18 of Translations of Mathematical Monographs. American Mathematical Society, Providence, R.I., 1969.
[12] Roger A. Horn and Charles R. Johnson. Matrix analysis. Cambridge University Press, Cambridge, second edition, 2013.
[13] Victor Kaftal and Gary Weiss. Traces on operator ideals and arithmetic means. J. Oper. Theory, 63(1):3–46, 2010.
[14] N. J. Kalton. Trace-class operators and commutators. J. Funct. Anal., 86(1):41–74, 1989.
[15] C. Laurie, E. Nordgren, H. Radjavi, and P. Rosenthal. On triangularization of algebras of operators. J. Reine Angew. Math., 327:143–155, 1981.
[16] Jiro Loreaux and Gary Weiss. Matrix splitting and ideals in \( B(H) \). Oper. Matrices, 2021. (to appear).
[17] Sasmita Patnaik. Ideals and Commutators of Operators. PhD thesis, University of Cincinnati, 2012. Thesis (Ph.D.)–University of Cincinnati.
[18] Sasmita Patnaik, Srdjan Petrovic, and Gary Weiss. Universal block tridiagonalization in \( B(H) \) and beyond. In The mathematical legacy of Victor Lomonosov. Operator theory, pages 317–326. Berlin: De Gruyter, 2020.
[19] Carl Pearcy and David Topping. On commutators in ideals of compact operators. Mich. Math. J., 18:247–252, 1971.
[20] R. Schatten. Norm ideals of completely continuous operators, volume 27. Springer-Verlag, Berlin, 1970.
[21] Gary Weiss. Commutators and Operator Ideals. PhD thesis, University of Michigan, 1975. Thesis (Ph.D.)–University of Michigan.
[22] Gary Weiss. Commutators of Hilbert–Schmidt operators. II. Integral Equations Operator Theory, 3(4):574–600, 1980.

\(^5\)The sequence \( s(A_n) \) is finite with length \( \min\{k_n, k_{n+1}\} \); but we may pad it with zeros to make it have length \( k_n \) in case \( k_{n+1} < k_n \).
