Abstract

Lin’s theorem states that for all $\epsilon > 0$, there is a $\delta > 0$ such that for all $n \geq 1$ if self-adjoint contractions $A, B \in M_n(\mathbb{C})$ satisfy $\|[A, B]\| \leq \delta$ then there are self-adjoint contractions $A', B' \in M_n(\mathbb{C})$ with $[A', B'] = 0$ and $\|A - A'\|, \|B - B'\| < \epsilon$. We present full details of the approach in [1], which seemingly is the closest result to a general constructive proof of Lin’s theorem. Constructive results for some special cases are presented along with applications to the problem of almost commuting matrices where $B$ is assumed to be normal and also to macroscopic observables.

1 Summary of Contents

This paper is concerning Lin’s theorem. We discuss this and related results, but are primarily concerned with putting down all the details of Matthew Hastings’ approach to Lin’s theorem in [1]. More precisely, we first discuss the history and related earlier results. We then discuss the earlier work of Davidson and Szarek from which it seems [1] builds. Some constructive results under certain conditions are discussed along the way. A goal of this paper is to add all the details in [1] and fix any errors. In particular, we factor out certain claims made in the paper, providing references or proofs for clarity. Calculations are intended to be in full (and perhaps too much) detail. For various modifications to [1], the author is indebted to Hastings for clarifications and resolving some errors. This is cited as [3].

The last three sections discuss constructibility and applications of this method to more general scenarios. An important note (discussed in an earlier section) is that Szarek’s paper already gives a constructive proof of a special (and still widely applicable) case of Lin’s theorem. The last section discusses Ogata’s result for macroscopic observables which can be seen as proving a positive result where Lin’s theorem does not generally apply. Some partial results and conjectures are discussed.

For a complete proof of Hastings’ result, see Sections 2 to 16 excluding Section 9. A short outline of the steps can be found in Section 3.

The following are some conventions used in this paper. Hilbert spaces are always finite dimensional and complex with an inner product $\langle -, - \rangle$ that is conjugate linear in

*Rutgers University. dh708@math.rutgers.edu
the first argument. \(\|\cdot\|\) always denotes the operator norm of a linear transformation on a Hilbert space (i.e. a matrix), \(|\cdot|\) always denotes the norm of a vector (or absolute value of a number), and \([C,D] = CD - DC\) is the commutator of two matrices. It is common in the literature on this problem to see self-adjoint matrices referred to as Hermitian. Often, this will be done in this paper when discussing the statements of Lin’s theorem or results from other papers. However, in the bulk of the paper, we will simply refer to them as “self-adjoint”.

2 Introduction

The following is known as Lin’s theorem:

For every \(\epsilon > 0\), there is a \(\delta > 0\) such that if \(A, B \in M_n(\mathbb{C})\) are self-adjoint contractions and \(\|[A,B]\| < \delta\), then there are commuting self-adjoint \(A', B' \in M_n(\mathbb{C})\) such that \(\|A - A'\|, \|B - B'\| < \epsilon\).

It is important to note that \(\delta\) does not depend on \(n\). We will discuss results that depend on \(n\) later.

The condition that \(\|[A,B]\|\) is small is often called “almost commuting” and the property that there are nearby commuting matrices (with whatever specified properties) is often called “nearly commuting”. In these terms, Lin’s theorem says that almost commuting Hermitian matrices are near Hermitian commuting matrices. Lin’s theorem can also be stated in terms of almost normal matrices being near normal matrices.

Lin’s theorem is the positive answer to a question (for the operator norm) stated by Rosenthal in [24] in 1969 and in 1976 was listed as one of various unsolved problems by Halmos in [23]. By the time that Lin’s proof appeared, there were various results that fell short of Lin’s theorem in important ways. Positive results were obtained when \(\delta\) is allowed to depend on \(n\) and negative results were obtained when we are considering general self-adjoint operators acting on an infinite dimensional Hilbert space. [25] in 1969 (with the Euclidean norm) and [26] in 1974 (with the operator norm, with proof attributed to Halmos) present a result with dimensional dependence of \(\delta\). Also, [26] presents related results that fail in infinite dimensions and [27] in 1980 presented an example of an obstruction related to the Fredholm index in the infinite dimensional analogue to Lin’s theorem. In 1979, Pearcy and Shields in [6] gave the dimensional dependence of \(\delta = 2\epsilon^2/(n - 1)\), but only used that one of the operators is assumed to be self-adjoint. The dependence of \(\delta\) on \(\epsilon\) has the optimal exponent for Lin’s theorem.

Davidson in [5] in 1985 provided an example of normal matrices (gotten by an approximation result of Berg in [31]) and a Hermitian matrices that almost commute but are not nearby commuting normal and Hermitian matrices using a spectral projection argument and a property of the shift on \(\mathbb{C}^n\). This shows that the Pearcy and Shields result cannot be improved to remove its dimensional dependence, as we will discuss later. See Section 18 for more about generalizations.

In 1990, Szarek in [17] proved that Lin’s theorem holds for \(\delta = c\epsilon^{13/2}/n^{1/2}\), where \(c\) is some constant. Szarek states that the exponent of \(\epsilon\) can be reduced by using a more complicated argument but the aim was to use the approach discussed in [5] and the assumption that both operators are Hermitian to obtain stronger dimensional dependence than the counter-examples seen thus far had exhibited and hence “the

\footnote{The dates stated in this paper use published dates and received dates when possible.}
situation in the Hermitian setting is completely different”.

In 1997, Lin in [1] provided a proof of Lin’s Theorem and in 1996 Friis and Rørdam in [16] published a simplified version of Lin’s result (referencing [4] which was to appear in Operator Algebras and Their Applications).

Up until this point, the proof of Lin’s theorem was abstract and neither gave the asymptotic dependence of $\delta$ on $\epsilon$ nor gave a way of constructing these matrices. In 2008, [2] by Hastings presented an approach whose starting point is similar to one of [5]’s reformulations of Lin’s theorem and [2] claimed to provide a constructive method of finding the nearby commuting matrices and also presenting a dependence of $\delta$ on $\epsilon$.

The methods in [2] for a diagonal and tridiagonal pair of Hermitian matrices (which was already implicitly solved without the ideal exponent in [17], as we will discuss later) held, but the arguments for the general case were amended to try to resolve errors and consequently the main result changed. Namely, the most recent version of this paper, being [1] posted to Arxiv.org in 2010, states similar statements for asymptotic dependence of $\delta$ on $\epsilon$ but not that the result is constructive and the dependence only holds for $\delta$ small enough (but neither how small nor any constant in the dependence of $\delta$ on $\epsilon$ is given).

A main purpose of this paper is to write out the details of a fully corrected version of [1], giving some different (at least differently stated) intuition for the proof, providing details and references, simplifying some arguments, and resolving some issues found by the author and resolved in communication with Hastings (cited in this paper as [3]). Some partial results that are either new or not previously put into writing in this way have also been collected in Section 9. We also provide some applications of this approach to constructibility of the nearby commuting matrices and the situation of a Hermitian matrix and a normal matrix.

A list of applications of almost/nearly commuting matrices can be found in the introduction of [8]. The first example given there is related to a paper of von Neumann concerning macroscopic observables. Although at least three almost commuting Hermitian matrices are not necessarily nearly commuting, Ogata in [7] showed that a special case of this related to von Neumann’s example is true. This is discussed in more detail in Section 19.

To continue the discussion of the history of the problem, there have been many “spin-offs” of this problem, involving different norms, algebraic objects that are not operators on a Hilbert space, and for bounded operators on an infinite dimensional Hilbert space. Notable mentions are that Lin’s theorem holds for the normalized Schatten-$p$ norms ([19]) and that Lin’s theorem for the normalized Hilbert-Schmidt norm holds for multiple Hermitian matrices ([11]). In 2015, Kachkovskiy and Safarov in [10] stated a result that not only gives the optimal dependence of $\delta$ on $\epsilon$ of $\epsilon \leq Const.\delta^{3/2}$ but simultaneously addresses the infinite dimensional situation where the index obstruction of Berg and Olsen applies. The proof does not appear to be constructive, but provides the optimal asymptotic dependence.

One of the equivalent forms of Lin’s theorem that Davidson uses is the following (we use notation from [11]):

\[(Q'):\text{For every } \epsilon > 0, \text{ there is a } L_0 > 0 \text{ with the following property. For any self-adjoint contraction } J \text{ that is block tridiagonal with } L \geq L_0 \text{ blocks, there is a projection } P \text{ that contains the first block of the basis, is orthogonal to the last block of the basis, and satisfies } \|[J,P]\| < \epsilon.\]
Moreover, a constructive proof of Lin’s theorem gives a constructive proof of \((Q')\) and vice-versa.

Although the optimal result for Lin’s theorem is gotten in \([10]\), there are other reasons to be interested in the methods of \([1]\) or an approach that uses \((Q')\). (For example, proving \((Q')\) as in Remark 7.5.)

First, one can use these estimates to get results for other types of matrices, like two unitary matrices with a spectral gap in \([11]\) or other situations like in \([9]\). We discuss a few of these in Section 18.

Second, the method in \([1]\) is to reduce the general case to one block tridiagonal matrix and one block identity matrix, then apply \((Q')\). Breaking these matrices up into consecutive blocks, by “pinching” \(^3\) one applies Lemma 2 of \([1]\) (or our Lemma 7.4 below) which gives a local joint near-diagonalization of the matrices. This could be interesting for applications of this result because one might hope that a basis on which both of the nearby commuting matrices are diagonal is localized with respect to a basis in which one of the original matrices is diagonal. See Section 19 for an application of this.

Third, this proof is the closest to a constructive proof that has been produced. The method in \([1]\) is not constructive as it utilizes Lin’s theorem in the above mentioned Lemma 2, which is our Lemma 7.1 The proof only uses Lin’s theorem for a specific value of \(\epsilon < 1/22\) so on its face seems weaker than Lin’s theorem. This is the bottleneck for a constructive proof of Lin’s theorem through \([1]\)’s method. An open question is if there is a constructive proof of this bottleneck lemma.

3 Outline of Hastings’ proof

The key result of \([1]\) that we will prove is of the form:

**Theorem 3.1.** For \(\delta > 0\) small enough there exists an \(\epsilon = \epsilon(\delta) > 0\) such that for any self-adjoint contractions \(A, B \in M_n(\mathbb{C})\) such that \(\|[A, B]\| \leq \delta\) there exist self-adjoint \(A', B'\) such that \([A', B'] = 0\) and \(\|A - A'\|, \|B - B'\| \leq \epsilon\). We can choose \(\epsilon(\delta) = E(1/\delta)^{\delta^\beta}, \beta > 0\) where \(E\) is a function growing slower than any polynomial.

The value of \(\beta\) gotten in \([1]\) depends on the type of matrix considered. Our exposition will prove this result with a different value for \(\beta\) than that of \([1]\) given some complications from its Lemma 4. This is due to a construction in Lemma 6 of \([1]\) that we simplify as suggested in \([3]\).

We start with two almost commuting self-adjoint contractions \(A, B \in M_n(\mathbb{C})\). We apply Lemma 6.1 to replace \(A\) with \(H\). \(H\) and \(B\) are almost commuting self-adjoint contractions and \(H\) has the additional property that it is “finite range” (we can choose how far the “range” \(\Delta\) is) with respect to the eigenspaces of \(B\) by which we mean that it only maps nearby eigenspaces of \(B\) into nearby eigenspaces. The price paid is that the distance between \(A\) and \(H\) will depend on how close \(A\) and \(B\) are to commuting and the range \(\Delta\).

Next, break the spectrum of \(B\) up into many small disjoint intervals \(I_i\). We choose the “finite range” to be much smaller than the length of the above intervals, so that

\(^2\)The blocks are composed of the same basis vectors.

\(^3\)If \(A\) is a matrix and we have orthogonal projections \(P_1, \ldots, P_k\) such that \(\sum_i P_i = I\) then \(\sum_i P_i A P_i\) is the “pinching” of \(A\) by \(\{P_i\}\) by the terminology in \([18]\).
each of these intervals breaks up into many subintervals of length at most $\Delta$. We form orthogonal subspaces by grouping the eigenvectors of $B$ associated with each subinterval. We project $H$ onto the eigenvectors of $B$ associated with the interval $I_i$ (the range of $E_{I_i}(B)$). Then on the range of each $E_{I_i}(B)$ the projection of $H$ will be tridiagonal with respect to the many subspaces associated with the subintervals. We apply $(Q')$, stated as Lemma 7.4 here to then obtain for each interval $I_i$ a subspace $W_i$ that is almost invariant under the projection of $H$, contains the eigenvectors associated with the first subinterval, and is orthogonal to the eigenvectors associated with the last subinterval. Each subspace has an orthogonal complement as a subspace of $E_{I_i}(B)$ which we call $W_i^\perp$. Now, we use these to form $A'$ by projecting $H$ onto the subspaces $W_i^\perp \oplus W_{i+1}$. We define $B'$ to be a multiple of the identity on $W_i^\perp \oplus W_{i+1}$. The details can be found in Section 7. The estimates for these various steps are collected in Section 10.

At this point, we would focus all of our attention on proving Lemma 7.1 which after a simple projection construction implies Lemma 7.4. The set-up is that have a self-adjoint contraction which is tridiagonal with respect to some orthogonal subspaces $\mathcal{V}_1, \ldots, \mathcal{V}_L$. In Section 8 there is a discussion of the general approach from both 11 and another relevant paper 17, which (among other things) is explored in Section 9.

A starting point is that the first two parts of Lemma 7.1 can be satisfied using a simple construction: break the spectrum of $J$ up into intervals $I_i$ (which are unrelated to the intervals discussed above), then constructions similar to $\mathcal{W} = \bigoplus X_i \omega$ with $X_i = \chi_{I_i}(J)(\mathcal{V}_1)$ are used to construct the subspace $\mathcal{W}$. The idea is that this subspace recovers $\mathcal{V}_1$ as a subspace and because each subspace is almost invariant under $J$ and they are orthogonal, the entire subspace is almost invariant under $J$. The key obstacle to be avoided is that we need $\mathcal{W}$ to be almost orthogonal to $\mathcal{V}_L$. The details and further motivation are in Section 8.

Section 11 lists some of the properties of the smooth cut-off functions $F_{\omega_i}(\omega)$ used and provides a definition of the key subspaces $X_i$. Section 13 discusses many of the subspaces derived from the $X_i$ and the first three paragraphs contains some of the general motivation in the construction of $\mathcal{W}$. An issue is that if $x = \sum x_i$ for $x_i \in X_i$, then we do not necessarily have control of $\sum |x_i|^2$ in terms of $|x|$. To approach this, we explore the subspace $X = \bigoplus X_i$ and representation space $R$ which is the exterior direct sum of the $X_i$. We form the linear map $A : R \to X$ that is the identity on each $X_i$ (which we call $R_i$ when it is a subspace of $R$) and we are interested in having some control of $\sum |r_i|^2$ in terms of $|x|^2 = |A\sum r_i|^2$.

In Section 14 we form the subspace $U_i$ from subspaces $N_i$ (i.e. even) and $N'_i$ (i.e. odd) which satisfy certain properties and set $\mathcal{W} = A(U)$. Roughly speaking, the subspaces $N_i$ belong to the span of the eigenvectors of $\rho = A^*A$ with “small eigenvalues” and can be put together to approximately recover in a reasonable way the eigenvectors of $\rho$ with small eigenvalues. This uses the tridiagonal nature of $\rho$, Lemma 6.10 and Lemma 12.1 (discussed later in this Section). For $i$ odd, we cut out the part of $X_i$ that is not orthogonal enough to the neighboring (even) $N_{i-1}$ and $N_{i+1}$ to obtain $N''_i$. This semi-orthogonality is used throughout the remainder of the proof.

In Section 15 some various inequalities concerning representation of vectors in $W$ are proved along with inequalities concerning the projection onto $W$. Finally, in Section 16 we conclude the proof with verifying the validity of the desired properties of $W$.

Section 4 (Linear Algebra Preliminaries), Section 5 (Lemmas on Spectral Projec-
tions and Commutators), and Section 6 (Relevant Lieb-Robinson Bounds) provide various important results that seem outside the “main story” and have been separated from the rest of the proof for clarity. Section 12 singles out Lemma 12.1, the “non-constructive bottleneck” of Hastings’ proof. We only use this lemma for one value of $\epsilon < 1$ in the construction of the $N_i$ to ensure that $\|N_i N_i N_{i+1}\| \leq 1 - \chi < 1$ for some $\chi > 0$. This involves applying Lin’s theorem for only one value of $\epsilon$ less than $1/22$. As described in [1], Hastings’ result bootstraps Lin’s theorem for one value of $\epsilon$ to get a result for an asymptotic dependence of $\delta$ on $\epsilon$.

4 Linear Algebra Preliminaries

Note that $| - |$ will always denote the (Hilbert space) norm of a vector and $\| - \|$ will always denote the operator norm of a matrix. If $S$ is a set and $N$ is normal, $E_S(N)$ is the spectral projection of $N$ on $S$. If $W$ is a subspace, $P_W$ is a projection onto $W$. Note that $S_1, S_2 \subset \mathbb{R}$ will always denote disjoint sets where the relevant constant is the distance between them and $S''$, $S' \subset \mathbb{R}$ denote nested sets and the relevant constant is the distance between $S''$ and $\mathbb{R} \setminus S'$. We will use the convention for the Fourier transform: $f_k = \frac{1}{2\pi} \int f(x)e^{-ikx}dx$.

Definition 4.1. If we have a sequence of vectors $u_k$ (resp. subspaces $U_k$) such that if $i, j$ with $|i - j| \geq 2$ then $u_i$ and $u_j$ (resp. $U_i$ and $U_j$) are orthogonal, we call this sequence nonconsecutively orthogonal.

One type of estimate used often in the proof of [1]’s Lemma 2 is that if $x_1, \ldots, x_n$ are nonconsecutively orthogonal vectors with sum $v$ then we have

$$|v|^2 = \left| \sum_i x_i \right|^2 = \left| \sum_{i \text{ odd}} x_i + \sum_{i \text{ even}} x_i \right|^2 \leq 2 \left| \sum_{i \text{ odd}} x_i \right|^2 + 2 \left| \sum_{i \text{ even}} x_i \right|^2 = 2 \sum_i |x_i|^2. \quad (1)$$

If one has uniform control of the inner products of vectors $x_i$ then one has a reverse inequality. Specifically, if there is also $C < 1/2$ such that

$$|(x_i, x_j)| \leq C|x_i||x_j| \quad (2)$$

for $j = i \pm 1$, we have

$$|v|^2 = \left| \sum_i x_i \right|^2 \geq \sum_i |x_i|^2 - 2 \sum_{i < j} |(x_i, x_j)| \geq \sum_i |x_i|^2 - 2C \sum_{1 \leq i < n} |x_i||x_{i+1}|$$

$$\geq \sum_i |x_i|^2 - 2C \sum_{1 \leq i < n} \frac{|x_i|^2 + |x_{i+1}|^2}{2} \geq (1 - 2C) \sum_i |x_i|^2. \quad (3)$$

This condition of some degree of minimal orthogonality is what makes Hastings’ tridiagonal result possible, because given any nonconsecutively orthogonal subspaces $X_k$ which are each one dimensional, either we have control on the inner products as in Equation (2) for each vector $x_i \in X_i, x_{i+1} \in X_{i+1}$ or we have the opposite inequality. This is the key element to the construction in Lemma 6 of [1].

We consider Proposition 2.2 from [32] which after some crude estimates, gives the next result. Recall that a matrix $A$ is called $m$-banded if $A_{i,j} = 0$ for $|i - j| > m/2$. Being 2-banded is equivalent to being tridiagonal.
Proposition 4.2. Let $A$ be an invertible tridiagonal positive definite matrix with $a = \min \sigma(A) > 0, b = \max \sigma(A)$. Then

$$|(A^{-1})_{i,j}| \leq C a^{|i-j|},$$

where $C = C(a, b) > 0, \alpha = \alpha(a, b) \in (0, 1)$.

Remark 4.3. Examples of results like this for analytic functions, instead of the function $f(x) = 1/x$, of banded self-adjoint matrices can be found in [15].

Lemma 4.4. Suppose that $c_i, d_i, i = 1, \ldots, n$ are non-negative constants. If $M$ is a self-adjoint $n \times n$ tridiagonal matrix such that $M_{i,i} \geq c_i^2 + d_i^2$ for $i \leq n$ and $|M_{i,i+1}| \leq d_i c_{i+1}$ for $1 \leq i < n$, then $M$ is positive.

Proof. We will compare it to the Hermitian tridiagonal matrix $D$ defined by $D_{i,i} = a_i^2 + b_i^2$ and $D_{i,i+1} = b_i \overline{a_{i+1}}$. Here is the case $n = 4$:

$$D = \begin{pmatrix}
|a_1|^2 + |b_1|^2 & \overline{b_1} a_2 & |a_2|^2 + |b_2|^2 & \overline{b_2} a_3 \\
|b_1| a_2 & |a_2|^2 + |b_2|^2 & \overline{b_2} a_3 & |a_3|^2 + |b_3|^2 \\
|b_2| a_3 & |a_3|^2 + |b_3|^2 & |a_3|^2 + |b_3|^2 \\
|b_3| a_4 & |a_4|^2 + |b_4|^2 & |b_3| a_4
\end{pmatrix}.$$

This matrix is positive because if we consider the matrix $G$ defined with columns $(a_1, b_1, 0, \ldots, 0)^T, (0, a_2, b_2, 0, \ldots, 0)^T, \ldots, (0, \ldots, 0, a_n)^T$ then we have that $G^*G + b_n^2 e^{n,n} = D$, where $(e^{n,n})_{i,j} = \delta^n_i \delta^n_j$. So, $D$ is positive.

When $a_i = c_i, b_i = d_i$, we see that the matrix $M$ is compared to $D$ by having its diagonal entries larger than those of $D$ and its off-diagonal entries smaller in absolute value than $D$. We now pick $a_i, b_i$ so that it is clear that $M$ is positive. Let $a_i = c_i$ for $i = 1, \ldots, n$. If $d_i c_{i+1} = 0$ then let $b_i = 0$. If $d_i c_{i+1} > 0$, let $b_i = M_{i,i+1}/\overline{a_{i+1}}$ for $i = 1, \ldots, n - 1$. This is so that $M_{i,i+1} = b_i \overline{a_{i+1}}$. Let $b_n = d_n$.

Now, $|b_i \overline{a_{i+1}}| \leq d_i c_{i+1}$ so $a_i = c_i$ and $|b_i| \leq d_i$. $|a_i|^2 + |b_i|^2 \leq c_i^2 + d_i^2 \leq M_{i,i}$. This tells us that $D$ has the same off-diagonal terms as $M$ and its diagonal terms are less than the diagonal terms of $M$. This tells us that $M - D$ is positive, so $M$ is positive.

We use Lemma 2 of [6] concerning the Schur product.

Lemma 4.5. Let $(T_{i,j})$ be a matrix and let $a_i, b_j$ be real numbers with $a_i - b_j \geq d$. Then

$$\left\| \left( \frac{1}{a_i - b_j} T_{i,j} \right) \right\| \leq \frac{1}{d} \| (T_{i,j}) \|.$$

The following (constructive) lemma which is Lemma 2.2 from [5], will serve many uses, including simplifying the statement of a main lemma in [1] and the lemma where Lin’s theorem is applied.

Lemma 4.6. Let $E \leq G$ be projections on a Hilbert space and $\epsilon > 0$. If $F'$ is a projection with $\|E F' + \| < \epsilon$ and $\|F' G - \| < \epsilon$, then there is a projection $F$ such that $E \leq F \leq G$ with $\|F - F'\| \leq 5\epsilon$. 

7
The following result concerning projections has been called “Jordan’s Lemma”. We restate the proof because it is simple and because we want to emphasize that the decomposition is orthogonal, because the results cited below that state the result in these terms do not clearly mention this property.

**Proposition 4.7.** Let $P, Q$ be two projections on finite dimensional Hilbert space $\mathcal{H}$. They induce an orthogonal decomposition of $\mathcal{H}$ into one and two dimensional spaces $\mathcal{H}_i$ that are invariant under both $P$ and $Q$ and irreducible in the sense that $P|_{\mathcal{H}_i} = Q|_{\mathcal{H}_i}$ only if $\dim \mathcal{H}_i = 1$.

**Remark 4.8.** Jordan’s lemma shows that orthogonal projections $P, Q$ onto subspaces in any dimension is just the direct sum of simple cases that we already understand: one dimensional projections in at most two dimensions. With this in mind, one sees parallels between Euler’s decomposition of rotation matrices in $\mathbb{R}^n$.

**Proof.** Consider the reflections $R = 2P - 1, S = 2Q - 1$ and the unitary operator $U = RS$. We only need to prove the result for $R$ and $S$, instead of $P$ and $Q$.

If $v$ is an eigenvector of $U$ with eigenvalue $\lambda$, then we claim that $H_1 = \text{span}\{v, Rv\}$ is invariant under $R$ and $S$. Clearly, it is invariant under $R = R^{-1}$ so we check invariance under $S$.

$$Sv = R^{-1}Uv = \lambda Rv.$$  

Also, $U^{-1} = S^{-1}R^{-1} = SR$ so

$$S(Rv) = U^*v = \bar{\lambda}v.$$  

So, $H_1$ is invariant under $R$ and $S$. This is a subspace of at most two dimensions and because $R$ and $S$ are self-adjoint we obtain that $H_1$ reduces $R$ and $S$. If $P$ and $Q$ agree on $H_1$, we can break down $H_1$ into one dimensional subspaces on which $P$ and $Q$ agree. Thus we can restrict $R$ and $S$ to the orthogonal complement of $H_1$ and the result then follows by infinite descent. 

A proof of this result and a discussion about this from the perspective of research in Quantum Computation can be found in Section 3.3.1 of [48]. Note that sometimes, as in Section 2.1 of [17] where the above proof is primarily taken, this result is stated in terms of unitary matrices with spectrum in $\{-1, 1\}$, which are just the reflection across the ranges of $P, Q$ given by $2P - 1, 2Q - 1$, as in [46]. Note that more “functional analytic” perspectives for results related to this can be found in [20], [21], and [22].

Note that there is no such generalization of Jordan’s lemma to more than two projections. In fact, Davis proved in [52] that the Banach algebra of all bounded linear operators on a separable Hilbert space is generated by only three projections (and the identity).

The way that we use Jordan’s lemma is the form from [1]:

**Proposition 4.9.** Let $P, Q$ be two projections on Hilbert space $\mathcal{H}$. Then there is a basis $\{p_i\}$ of the range of $P$ such that $(p_i, Qp_j) = 0$ for $i \neq j$.

**Proof.** Let $\mathcal{H} = \sum_i \mathcal{H}_i$ as in Proposition 4.7 just above. If $\mathcal{H}_i$ is one dimensional, then it is an eigenspace for both $P$ and $Q$, so if it is a 1-eigenspace of $P$ let $p_i$ be a unit vector spanning $\mathcal{H}_i$, otherwise we do nothing. If $\mathcal{H}_i$ is two dimensional, then because
$H_i$ is invariant under $P$, it has an eigenvector there. Because $P$ is not a multiple of the identity when restricted to $H_i$, we obtain a 1-eigenvector $p_i$ for $P$ which spans the image of $P$ restricted to $H_i$.

We obtain that the span of the $p_i$ is the range of $P$ and because the $H_i$ are orthogonal, the $QP_i \in H_i$ are as well. So, the $QP_i$ are orthogonal and so $(p_i, Qp_j) = 0$ for $i \neq j$.

**Remark 4.10.** This result has the following geometric interpretation.

If $P \leq Q$, then we can pick any basis $\{p_i\}$ of the range of $P$. If $Q \leq P$, then we can form $\{p_i\}$ as an orthonormal basis of the range of $Q$ and extend it to an orthonormal basis of the range of $P$.

If $P \not\leq Q$, then there is an annoying fact that it may be true that two vectors $v, w$ in the range of $P$ may be orthogonal, but $Qv$ and $Qw$ may not be. For example, let $H = \mathbb{C}^3$ and let $P$ project onto the subspace spanned by the first two standard basis vectors $e_1, e_2$. If the range of $Q$ is the span of $e_3$ and $e_1 + e_2$, then $Qe_1 = Qe_2$. In other words, $e_1, e_2$ are orthogonal but by applying $Q$ we have eliminated the components of $e_1$ and $e_2$ that contribute to their orthogonality. A way to avoid this phenomenon is to pick basis vectors $v = \frac{1}{\sqrt{2}}(e_1 + e_2)$ and $w = \frac{1}{\sqrt{2}}(e_1 - e_2)$ for the range of $P$ so that $Qv = v$ and $Qw = 0$, so that $\{Qv, Qw\}$ is an orthogonal set of vectors.

This construction is more complicated in the general case when $P$ and $Q$ do not intersect orthogonally or when there are multiple two dimensional subspaces in the decomposition. In particular, even though $\{QP_i\}$ are orthogonal, we are not guaranteed anything about the norm of these vectors. This is easily seen in the case that $P$ and $Q$ project onto arbitrary lines in $\mathbb{C}^2$.

5 **Lemmas on Spectral Projections and Commutators**

**Proposition 5.1.** (Davis-Kahan $\sin \theta$ Theorem) There exists a constant $c > 0$ such that for self-adjoint $A, B \in M_n(\mathbb{C})$ and $S_1, S_2 \subset \mathbb{R}$ we have

$$\|E_{S_1}(A)E_{S_2}(B)\| \leq \frac{c}{\text{dist}(S_1, S_2)}\|A - B\|.$$  

If there is a $\delta > 0, \alpha, \beta \in \mathbb{R}$ with $S_1 \subset [\alpha, \beta]$ and $S_2 \subset (-\infty, \alpha - \delta] \cup [\beta + \delta, \infty)$, we have

$$\|E_{S_1}(A)E_{S_2}(B)\| \leq \frac{1}{\delta}\|A - B\|.$$  

**Proof.** For a proof see Sections 10 and 11 of [39].

**Remark 5.2.** If $P, Q$ are projections, the quantity $\|PQ\|$ can be thought of as the “minimal $|\cos \theta|$” between any two lines in the range of $P$ and $Q$, respectively, because

$$\|PQ\| = \max_{|v|, |w| = 1} |(v, PQw)| = \max_{|v| = 1, Pw = v, Qw = w} |(v, w)|.$$  

With Jordan’s lemma if $\dim H_i = 2$ for all $i$, this gives $\|PQ\| = \max_i \cos \theta_i$, where $\theta_i$ are the angles between the rank one projections of $P$ and $Q$ restricted to $H_i$.  


If there are one dimensional \( \mathcal{H}_i \), then \( \|PQ\| = \max(\max_i \cos \theta_i, \max_j \|PQ|_{\mathcal{H}_j}\|) \), where \( i \) ranges over the two dimensional subspaces and \( j \) ranges over the one dimensional subspaces. \( \|PQ|_{\mathcal{H}_j}\| \) equals zero if \( P|_{\mathcal{H}_j} \) and \( Q|_{\mathcal{H}_j} \) are not identical and equals one otherwise.

In our notation, [39] states that “the name ‘\( \sin \theta \) theorem’ comes from the interpretation of \( \|PQ\| \) as the sine of the angle between \( \text{Ran}(P) \) and \( \text{Ran}(Q^\perp) \).”

**Example 5.3.** Consider \( A = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \), \( A_1 = \begin{pmatrix} 1 + \epsilon_1 & 0 \\ 0 & 0 \end{pmatrix} \), \( A_2 = \begin{pmatrix} 1 & \epsilon_2 \\ 0 & 0 \end{pmatrix} \). Perturbing \( A \) to get \( A_1 \) the eigenvalues (but not eigenvectors) to drift with \( \|A - A_1\| = \epsilon_1 \) and \( E_{\{1\}}(A) = E_{\{1+\epsilon_1\}}(A_1) \). For \( A_2 \), we get that the eigenvectors rotate but the eigenvalues remain unchanged.

(See [18] for more about this behavior in general.)

Here are two results regarding spectral projections and commutators. One result has the commutator small and the other has that the operators have a small difference. This result is part of an argument used in [6].

**Proposition 5.4.** Suppose that \( C, D \in M_n(\mathbb{C}) \) with \( C, D \) self-adjoint. Then for \( S_1, S_2 \subset \mathbb{R} \), we have

\[
\|E_{S_1}(D)CE_{S_2}(D)\| \leq \frac{\|C, D\|}{\text{dist}(S_1, S_2)}.
\]

**Proof.** Fix a vector \( v \) in the range of \( E_{S_1}(D) \) and a vector \( w \) in the range of \( E_{S_2}(D) \). Because \( v, w \) are arbitrary, we wish to show that

\[
|\langle v, Cw \rangle| \leq \frac{1}{\text{dist}(S_1, S_2)} \|\langle C, D\|\|.
\]

Let the notation: \( v_\lambda \) (which may be zero) represent a vector such that \( Dv_\lambda = \lambda v_\lambda \) and \( w_\mu \) likewise. That is, \( v_\lambda, w_\mu \) are eigenvectors or zero. We write the orthogonal eigenspace decompositions \( v = \sum_\lambda v_\lambda, w = \sum_\mu w_\mu \), where for the rest of the proof \( \lambda \) will be an element of \( S_1 \) and \( \mu \) an element of \( S_2 \).

Because \( \lambda, \mu \in \mathbb{R} \) and \( D^* = D \), we see that

\[
(\lambda - \mu)(v_\lambda, Cw_\mu) = (Dv_\lambda, Cw_\mu) - (v_\lambda, CDw_\mu) = -(v_\lambda, [C, D]w_\mu).
\]

List the eigenvectors \( v_\lambda \neq 0 \) as \( u_1, \ldots, u_r \) and the \( \lambda \)'s as \( a_1, \ldots, a_r \). Also list the eigenvectors \( w_\mu \neq 0 \) as \( u_1, \ldots, u_n \) and the \( \mu \)'s as \( b_s, \ldots, b_n \). Let \( u_{r+1}, \ldots, u_{s-1} \) be some unit vectors so that \( B = \left\{ \frac{u_i}{\|u_i\|} \right\} \) forms an orthonormal basis for \( \mathbb{C}^n \). Define otherwise \( a_i = 0, u_i = \text{dist}(S_1, S_2) \). We define a matrix \( T \in M_n(\mathbb{C}) \) in the basis \( B \) so that if \( 1 \leq i \leq r \) and \( s \leq j \leq n \) then \( T_{i,j} = \left( \frac{u_i}{\|u_i\|}, [C, D]u_j \right) \) and \( T_{i,j} = 0 \) otherwise. This is so that our auxiliary operator \( T \) satisfies \( (u_i, Tuj) = (u_i, [C, D]u_j) \) when \( 1 \leq i \leq r \) and \( s \leq j \leq n \) and \( T = E_{S_1}(D)[C, D]E_{S_2}(D) \).

Applying Lemma 4.5 we obtain

\[
|\langle v, Cw \rangle| = \sum_{\lambda, \mu} \frac{1}{\lambda - \mu} (v_\lambda, [C, D]w_\mu) = \sum_{\lambda, \mu} \frac{1}{\lambda - \mu} (v_\lambda, Tw_\mu)
\]

\[
= \sum_{i,j} \frac{1}{a_i - b_j} (u_i, Tuj) = \langle v, \left( \frac{1}{a_i - b_j} T_{i,j} \right) w \rangle \leq \frac{1}{\text{dist}(S_1, S_2)} \|\langle C, D\|\|.
\]

\( \square \)
Example 5.5. This result is nicely illustrated by taking $D = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}$ and $C = \begin{pmatrix} 0 & \epsilon \\ \epsilon & 0 \end{pmatrix}$. Then $[C, D] = \begin{pmatrix} 0 & -(b-a)\epsilon \\ (b-a)\epsilon & 0 \end{pmatrix}$ and $E_{\{a\}}(D)CE_{\{b\}}(D) = \epsilon I$.

This example shows that the result is sharp. We also see the behavior that if $C$ and $D$ almost commute and if $b-a$ is large then $\epsilon$ is small. Alternatively, if $b-a$ is small then $\epsilon$ is not required to be small (but it cannot be large). This can be interpreted as saying that $C$ approximately does not send vectors in one eigenspace of $D$ into a “far away” eigenspace of $D$ but can for nearby eigenspaces.

The following is a simplification of Theorem 3.2.32 from [33] that has been modified so that it agrees with our convention of the Fourier transform from [1].

Proposition 5.6. Let $f \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$ with $Cf = \int_\mathbb{R} |k| |\hat{f}(k)| dk < \infty$. (4)

If $A, B \in M_n(\mathbb{C})$ with $A$ self-adjoint, then

$$f(A) = \int_\mathbb{R} \hat{f}(k)e^{ikA} dk,$$

$$[f(A), B] = i \int_\mathbb{R} k\hat{f}(k) \int_0^1 e^{itkA}[A, B]e^{i(1-t)kA} dt dk$$

and so

$$\|[f(A), B]\| \leq C_f\|[A, B]\|.$$

We apply Proposition 5.6 to get the following two results:

Lemma 5.7. If $A, B \in M_n(\mathbb{C})$ with $A$ self-adjoint and having no spectrum in $(a, b)$, then

$$\|[E_{(-\infty, a]}(A), B]\| \leq C_2 \frac{b-a}{c_2} \|[A, B]\|.$$

Here $c_2 = \frac{2}{\pi} \inf \|\hat{\rho}\|_{L^1(\mathbb{R})}$, where the infimum is taken over all $\rho \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$ supported in $[-1, 1]$ with $\int_\mathbb{R} \rho = 1$.

Proof. Write $b = a + 2\epsilon_0$. We restrict to $0 < \epsilon < \epsilon_0$. Let $f_\epsilon(x) = \chi_{[R-\epsilon, a+\epsilon]} * \rho_\epsilon$, where $\rho$ is as above, $\rho_\epsilon(x) = \frac{1}{\epsilon} \rho(x/\epsilon)$, and $R = \min \sigma(A)$. Then $f_\epsilon(x)$ equals 1 on $[R, a]$ and equals zero outside $(-R - 2\epsilon, b)$. Because $A$ has no spectrum in $(-\infty, R) \cup (a, b)$, we see that $f_\epsilon(A) = E_{(-\infty, a]}(A)$. Recalling that

$$\hat{\chi}_{[\epsilon, a]}(k) = e^{-ik\frac{a+\epsilon}{2}} \sin(\frac{d-\epsilon}{2} k) \frac{1}{\pi k},$$

we obtain

$$C_{f_\epsilon} = \int_\mathbb{R} |k\hat{\chi}_{[R-\epsilon, a+\epsilon]}(k)\hat{\rho}(ek)| dk \leq \frac{1}{\pi \epsilon} \int_\mathbb{R} |\hat{\rho}(ek)| \epsilon dk = \frac{1}{\pi \epsilon} \|\hat{\rho}\|_{L^1(\mathbb{R})}$$

and the result follows. \qed
Remark 5.8. There are other ways to pick the interpolating function $f_k$ in the proof, but ultimately we know that this result, up to the constant, is sharp and the best constant is at least 1. This is because
\[
\|E_{(-\infty,a)}(A)B - BE_{(-\infty,a)}(A)\| \geq \|E_{(-\infty,a)}(A)B - BE_{(-\infty,a)}(A)\| = \|(1 - E_{(-\infty,a)}(A))BE_{(-\infty,a)}(A)\| = \|E_{[b,\infty)}(A)BE_{(-\infty,a)}(A)\|
\]
and we know that we have the strict estimate $\|E_{[b,\infty)}(A)BE_{(-\infty,a)}(A)\| \leq c c_{\frac{1}{\Delta}} \|[A,B]\|$ from Example 5.5.

6 Relevant Lieb-Robinson Bounds

The following result appeared in [5] in the discussion following its Lemma 3.1, while the statement and proof appearing here is modified from [1].

Lemma 6.1. There exist constants $c_0, c_1 > 0$ such that given $\Delta > 0$ and self-adjoint $A, B \in M_n(\mathbb{C})$, there exists a self-adjoint $H$ such that
\[
\|A - H\| \leq \frac{c_0}{\Delta} \|[A,B]\|,
\]
\[
\|[H,B]\| \leq c_1 \|[A,B]\|,
\]
and $E_{S_1}(B)HE_{S_2}(B) = 0$ if $\text{dist}(S_1, S_2) \geq \Delta$.

Proof. Let $f \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$ be supported in $[-1,1]$ with $f(0) = 1$ such that the constants $c_0, c_1$ defined below are finite. Write
\[
H = \int_{\mathbb{R}} e^{i \frac{k}{\Delta}} A e^{-i \frac{k}{\Delta}} \hat{f}(k) dk.
\]

To show that $E_{S_1}(B)HE_{S_2}(B) = 0$ if $\text{dist}(S_1, S_2) \geq \Delta$, we pick $v_\lambda, v_\mu$ two eigenvectors of $B$ with $\lambda \in S_1, \mu \in S_2$. Then
\[
(v_\lambda, Hv_\mu) = \int_{\mathbb{R}} \left( e^{-i \frac{k}{\Delta}} v_\lambda , A e^{-i \frac{k}{\Delta}} v_\mu \right) \hat{f}(k) dk = (v_\lambda, Av_\mu) \int_{\mathbb{R}} e^{i \frac{\lambda - \mu}{\Delta} k} \hat{f}(k) dk = (v_\lambda, Av_\mu) f \left( \frac{\lambda - \mu}{\Delta} \right) = 0.
\]

A key fact (see [33] Lemma 3.2.31) is that for $k \in \mathbb{R}$, $\|[A,e^{ikB}]\| \leq |k| \|[A,B]\|$. We then see that because $f(0) = 1$,
\[
\|A - H\| = \left\| A \int_{\mathbb{R}} e^{i \frac{k}{\Delta} B} e^{-i \frac{k}{\Delta} B} \hat{f}(k) dk - \int_{\mathbb{R}} e^{i \frac{k}{\Delta} B} A e^{-i \frac{k}{\Delta} B} \hat{f}(k) dk \right\|
\]
\[
= \left\| \int_{\mathbb{R}} [A, e^{i \frac{k}{\Delta} B}] e^{-i \frac{k}{\Delta} B} \hat{f}(k) dk \right\| \leq \frac{\|[A,B]\|}{\Delta} \int_{\mathbb{R}} |k| \hat{f}(k) dk = \frac{c_0}{\Delta} \|[A,B]\|,
\]
where $c_0 = \int_{\mathbb{R}} |k| \hat{f}(k) dk$. Also,
\[
\|[H,B]\| = \left\| \int_{\mathbb{R}} e^{i \frac{k}{\Delta} B} [A,B] e^{-i \frac{k}{\Delta} B} \hat{f}(k) dk \right\| \leq \|[A,B]\| \int_{\mathbb{R}} |\hat{f}(k) dk = c_1 \|[A,B]\|,
\]
where $c_1 = \int_{\mathbb{R}} |f(k) dk$. \[\Box\]
Remark 6.2. The “best choice” of the constants $c_0, c_1$ depends on the function $f$ that we chose. A similar remark concerning the constant in Proposition 5.1 can be made where the geometry of $S_1$ and $S_2$ are more general. See [39] and [40].

It is stated in [1] that $c_1$ can be chosen to equal 1 with the provided function $f(x) = (1 - x^2)^3 X_{[-1,1]}(x)$, but this is not so because $f(0) > 0, f(10) < 0$ so $\|f\|_{L^1} > f(0) = 1$. In [5], Davidson uses a very similar but less direct proof for this result (saying that it is a modification of Theorem 4.1 in [38]) and by citing some literature obtains $c_0 = 8, c_1 = 4$.

We also have the following which generalizes the above.

Lemma 6.3. Let $c_0, c_1$ be the constants from Lemma 6.1. For $\Delta > 0$, self-adjoint $A \in M_n(\mathbb{C})$, and commuting self-adjoint $B_1, \ldots, B_m \in M_n(\mathbb{C})$, there exists a self-adjoint $H$ such

$$\|A - H\| \leq \frac{c_0 c_1^{m-1} \Delta}{\Delta} \sum_{j=1}^m \|[A, B_j]\|,$$

$$\|[H, B_j]\| \leq c_1^m \|[A, B_j]\|,$$

and $E_{S_1}(B_j)HE_{S_2}(B_j) = 0$ if $\text{dist}(S_1, S_2) \geq \Delta$.

Proof. We essentially iterate the above construction because the $B_j$ commute. Let $f$, $c_0$, and $c_1$ be as in the proof of Lemma 6.1 and set

$$H = \int_{\mathbb{R}^m} e^{i \sum_{j=1}^m \frac{k_j}{\Delta} B_j} A e^{-i \sum_{j=1}^m \frac{k_j}{\Delta} B_j} \hat{f}(k_1) \cdots \hat{f}(k_m) dk_1 \cdots dk_m.$$

Following the calculation in the argument in the proof of Lemma 6.1 and that the $B_j$ commute, that $[A, -]$ is a derivation, and $f(0) = 1$,

$$\|A - H\| = \left\| \int_{\mathbb{R}^m} \left[ A, \prod_{j=1}^m e^{\frac{k_j}{\Delta} B_j} \right] e^{-i \sum_{j=1}^m \frac{k_j}{\Delta} B_j} \hat{f}(k_1) \cdots \hat{f}(k_m) dk_1 \cdots dk_m \right\|$$

$$\leq \sum_{j=1}^m \|[A, B_j]\| \int_{\mathbb{R}^m} \frac{|k_j|}{\Delta} |\hat{f}(k_1)| \cdots |\hat{f}(k_m)| dk_1 \cdots dk_m$$

$$= \frac{c_0 c_1^{m-1} \Delta}{\Delta} \sum_{j=1}^m \|[A, B_j]\|$$

and

$$\|[H, B_{j_0}]\| = \left\| \int_{\mathbb{R}^m} e^{i \sum_{j=1}^m \frac{k_j}{\Delta} B_j} [A, B_{j_0}] e^{-i \sum_{j=1}^m \frac{k_j}{\Delta} B_j} \hat{f}(k_1) \cdots \hat{f}(k_m) dk_1 \cdots dk_m \right\|$$

$$\leq c_1^m \|[A, B_{j_0}]\|.$$
Using the last result of Lemma 6.3 we obtain $(v, H w)$ where $\Re \lambda$ bounded above by $\sqrt{\Delta}$. These two inequalities give the two inequalities of the lemma.

For $N$ normal, we can write it as the sum of the commuting $\frac{N + N^*}{2}$, $i \frac{N - N^*}{2i}$ and so we have the following consequence.

**Corollary 6.4.** Let $c_0, c_1$ be the constants from Lemma 6.1. For $\Delta > 0$, self-adjoint $A$, $N$ normal in $M_n(\mathbb{C})$, there exists a self-adjoint $H$ such that

$$\|A - H\| \leq \frac{2c_0 c_1}{\Delta} \|[A, N]\|,$$

$$\|[H, N]\| \leq 2c_2 \|[A, N]\|,$$

and $E_{S_1}(N) HE_{S_2}(N) = 0$ if $S_1, S_2 \subset \mathbb{C}$ and $\text{dist}(S_1, S_2) \geq \sqrt{2} \Delta$.

**Proof.** Apply Lemma 6.3 with $B_1 = \Re N = \frac{N + N^*}{2}$, $B_2 = \Im N = \frac{N - N^*}{2i}$ using

$$\|[A, B_j]\| \leq \|[A, N]\|,$$

$$\|[H, N]\| \leq \|[H, \Re N]\| + \|[H, \Im N]\|.$$

These two inequalities give the two inequalities of the lemma.

To obtain the third result, note that the distance between points $\lambda \in S_1, \mu \in S_2$ is bounded above by

$$\sqrt{2} \max(|\lambda_1 - \mu_1|, |\lambda_2 - \mu_2|),$$

where $\Re \lambda = \lambda_1$, $\Im \lambda = \lambda_2$, $\Re \mu = \mu_1$, $\Im \mu = \mu_2$. So, if $v$ is a $\lambda$-eigenvector for $N$ and $w$ is a $\mu$-eigenvector for $N$ with $\lambda \in S_1, \mu \in S_2$ then it is a $\lambda_j$-eigenvector for $B_j$ and $w$ is a $\mu_j$-eigenvector for $B_j$ and

$$\sqrt{2} \Delta \leq |\lambda - \mu| \leq \sqrt{2} \max(|\lambda_1 - \mu_1|, |\lambda_2 - \mu_2|).$$

Using the last result of Lemma 6.3, we obtain $(v, H w) = 0$.

The following Lieb-Robinson type result is from [1] where its statement and proof originate.

**Theorem 6.5.** Let $H, B$ self-adjoint be such that $\|H\| \leq 1$ and $E_{S_1}(B) HE_{S_2}(B) = 0$ if $\text{dist}(S_1, S_2) \geq \Delta$ for $S_1, S_2 \subset \mathbb{R}$. Let $v_{LR} = e^{2 \Delta}$. Then for $|t| \leq \text{dist}(S_1, S_2)/v_{LR}$,

$$\|E_{S_1}(B)e^{itH} E_{S_2}(B)\| \leq e^{-\text{dist}(S_1, S_2)/\Delta}. \quad (5)$$

**Proof.** By iteration we get that the range of $H^n E_{S_2}(B)$ lies in the range of spectral projection of $B$ on the open $n \Delta$ neighborhood of $S_2$. So, $E_{S_1}(B) H^n E_{S_2}(B) = 0$ if $\text{dist}(S_1, S_2) \geq n \Delta$. Then expressing $e^{itH}$ in $E_{S_1}(B)e^{itH} E_{S_2}(B)$ as the standard exponential power series, we get for $m = \lceil \frac{\text{dist}(S_1, S_2)}{\Delta} \rceil$

$$\|E_{S_1}(B)e^{itH} E_{S_2}(B)\| \leq \sum_{n \geq m} \frac{|t|^n}{n!} \leq \frac{1}{e} \sum_{n \geq m} \left( \frac{|t|}{n} \right)^n \leq \frac{1}{e} \left( \frac{e|t|/m}{1 - e|t|/m} \right)^m.$$
where the second inequality follows from the following reductions. \( n! \geq e(n/e)^n \) follows from the inequality

\[
\frac{\log(n!)}{n} \geq \log \left( \frac{n(n+1)}{2n} \right) \geq \log \left( \frac{n}{e^{1-1/n}} \right).
\]

The first equality is a convexity inequality and the second inequality follows by removing log’s giving \( n(1/2 - 1/e^{1-1/n}) + 1/2 \geq 0 \). A computation verifies the cases \( n = 1, 2, 3 \) and when \( n > 1/(1 - \log(2)) \approx 3.26 \), we have \( 1/2 - 1/e^{1-1/n} > 0 \).

Then because \( e|t|/m \leq \frac{\text{dist}(S_1, S_2)}{em\Delta} \leq e^{-1} \) we have

\[
\| E_{S_1}(B)e^{itH}E_{S_2}(B) \| \leq e^{-m}/(1 - e^{-1}) = e^{-m}/(e - 1) \leq e^{-\frac{\text{dist}(S_1, S_2)}{\Delta}}.
\]

\( \square \)

**Remark 6.6.** Note that this does not actually use the fact that \( H \) is self-adjoint or that \( t \) is real, but the power series expression of \( f(x) = e^{itx} \) applied to the matrix \( H \) with norm at most 1. Likewise, one might expect there to be similar estimates for other analytic functions \( f \). See the reference in Remark 4.3.

This is in line with the interpretation of Theorem 6.5 in terms of the coefficients of a matrix as follows. Let \( \beta = (v_1, \ldots, v_n) \) be some basis of \( \mathbb{C}^n \) and \( Bv_j = jv_j \), be a “position” operator which scales each basis vector by its index. Then the condition \( E_{S_1}(B)HE_{S_2}(B) = 0 \) for \( \text{dist}(S_1, S_2) \geq \Delta \) tells us that \( H \) is \( 2[\Delta] \)-banded. Then, as mentioned in Remark 4.3, one expects exponential decay of the entries away from the diagonal for analytic \( f \) (with some conditions). This is similar to what we have in the next result, which is key to our application of the Lieb-Robinson result. More generally, we will not look at \( m \)-banded matrices, but block tridiagonal matrices.

**Corollary 6.7.** Let \( H, B \) self-adjoint be such that \( \| H \| \leq 1 \) and \( E_{S_1}(B)HE_{S_2}(B) = 0 \) if \( \text{dist}(S_1, S_2) \geq \Delta \). Then for \( f \in C^0(\mathbb{R}) \cap L^1(\mathbb{R}) \),

\[
\| E_{S_1}(B)f(H)E_{S_2}(B) \| \leq \int_{|k| > \frac{\text{dist}(S_1, S_2)}{\epsilon \Delta}} |\hat{f}(k)|dk + \| \hat{f} \|_{L^1(\mathbb{R})}e^{-\frac{\text{dist}(S_1, S_2)}{\Delta}}.
\]

**Proof.** By the representation result in Proposition 5.6, and by Theorem 6.5, we see that

\[
\| E_{S_1}(B)f(H)E_{S_2}(B) \| \leq \int_{\mathbb{R}} \| E_{S_1}(B)e^{ikH}E_{S_2}(B) \| |\hat{f}(k)|dk
\]

\[
\leq \int_{|k| > \frac{\text{dist}(S_1, S_2)}{\epsilon \Delta}} |\hat{f}(k)|dk + \int_{|k| \leq \frac{\text{dist}(S_1, S_2)}{\epsilon \Delta}} e^{-\frac{\text{dist}(S_1, S_2)}{\Delta}} |\hat{f}(k)|dk
\]

\[
\leq \int_{|k| > \frac{\text{dist}(S_1, S_2)}{\epsilon \Delta}} |\hat{f}(k)|dk + \| \hat{f} \|_{L^1(\mathbb{R})}e^{-\frac{\text{dist}(S_1, S_2)}{\Delta}}.
\]

\( \square \)

**Remark 6.8.** We will use the smoothness of the function \( f \) (or functions), so that \( \Phi(t) = \int_{|k| > 1} |\hat{f}(k)|dk \) decreases faster than any polynomial and \( \| \hat{f} \|_{L^1} \) is at most a constant, to obtain fast decay of \( \| \hat{f} \|_{L^1} \). Note that if we do not care for “faster than any polynomial”, then we could just assume some smoothness for \( f \). See the end of Section 11 for some estimates.
The above two statements are more-or-less explicit in [1]. The following are implicitly used and, along with the former results, go under the umbrella of “Lieb-Robinson bounds”:

**Theorem 6.9.** Let $H, B$ self-adjoint be such that $\|H\| \leq 1$ and $E_{S'}(B)HE_{S''}(B) = 0$ if $S'' \subset S'$, $\text{dist}(S'', \mathbb{R} \setminus S') \geq \Delta$ for $S', S'' \subset \mathbb{R}$. Let $\nu_{LR} = e^{2\Delta}$. Then for $|t| \leq \text{dist}(S'', \mathbb{R} \setminus S')/\nu_{LR}$,

$$\| e^{itH} - e^{itE_{S'}(B)HE_{S''}(B)} \| E_{S''}(B) \| \leq 3e^{-\text{dist}(S'', \mathbb{R} \setminus S')/\Delta}. \quad (7)$$

**Proof.** The proof proceeds essentially as that of the “original” Lieb-Robinson result, Theorem 6.5, taking $S_1 = \mathbb{R} \setminus S'$ and $S_2 = S''$. For $n > 0$, let $S''_n$ be the open $n\Delta$ neighborhood of $S''$ and $S'_n = S''_n$.

Just as in the proof of Theorem 6.5 $H$ maps the range of $E_{S''_n}(B)$ into the range of $E_{S'_n}(B)$ and hence $H^n$ maps the range of $E_{S''}(B)$ into the range of $E_{S'_n}(B)$. We prove by induction that

$$[E_{S'}(B)HE_{S''}(B)]^n E_{S''}(B) = E_{S'}(B)H^n E_{S''}(B)$$

for $n \leq m = \left\lceil \frac{\text{dist}(S_1, S_2)}{\Delta} \right\rceil$ by noting it is clearly true when $n = 0, 1$ and that when $n \leq m - 1$, $S''_n \subset S'$ and hence

$$[E_{S'}(B)HE_{S''}(B)]^{n+1} E_{S''}(B) = [E_{S'}(B)HE_{S''}(B)][E_{S'}(B)H^n E_{S''}(B)] E_{S''}(B).$$

Because $H^n$ maps the range of $E_{S''}(B)$ into the range of $E_{S''_n}(B) \leq E_{S'}(B)$ so

$$[E_{S'}(B)HE_{S''}(B)][E_{S'}(B)H^n E_{S''}(B)] E_{S''}(B) = [E_{S'}(B)HE_{S''}(B)] E_{S''_n}(B)H^n E_{S''}(B) = E_{S'}(B)HE_{S''_n}(B)H^n E_{S''}(B) = E_{S'}(B)H^{n+1} E_{S''}(B).$$

This gives us the desired result.

Expressing the exponentials in $E_{S'}(B) \left[ e^{itH} - e^{itE_{S'}(B)HE_{S''}(B)} \right] E_{S''}(B)$ as the standard power series, we obtain

$$\| E_{S'}(B) \left[ e^{itH} - e^{itE_{S'}(B)HE_{S''}(B)} \right] E_{S''}(B) \| = \left\| E_{S'}(B) \left[ \sum_{n=0}^{\infty} \frac{(it)^n H^n}{n!} - \sum_{n=0}^{\infty} \frac{(it)^n (E_{S'}(B)HE_{S''}(B))^n}{n!} \right] E_{S''}(B) \right\| \\
\leq 2 \sum_{n \geq \| S'' \cap \mathbb{R} \setminus S' \|} \left\| \frac{|t|^n}{n!} \leq 2e^{-\text{dist}(S'', \mathbb{R} \setminus S')/\Delta}, \right\|$$

where the last inequality follows by the argument in the proof of Theorem 6.5. Also, by Theorem 6.5

$$\| E_{S'}(B)e^{itH} E_{S''}(B) - e^{itH} E_{S''}(B) \| = \| E_{\mathbb{R} \setminus S'}(B)e^{itH} E_{S''}(B) \| \leq e^{-\text{dist}(S'', \mathbb{R} \setminus S')/\Delta}.$$

The result then follows because, expanding as a power series, we see that

$$E_{S'}(B)e^{itE_{S'}(B)HE_{S''}(B)} E_{S''}(B) = e^{itE_{S'}(B)HE_{S''}(B)} E_{S''}(B).$$
Corollary 6.10. Let $H, B$ self-adjoint be such that $\|H\| \leq 1$ and $E_{S'}(B)HE_{S'}(B) = 0$ if $\text{dist}(S'' \setminus S') \geq \Delta$. Then for $f \in C^0(\mathbb{R}) \cap L^1(\mathbb{R})$, let $H' = E_{S'}(B)HE_{S'}(B)$ so

$$\|f(H) - f(H')E_{S'}(B)\| \leq 2 \int_{|t| > \frac{\text{dist}(S'' \setminus S')}{e^2\Delta}} |\hat{f}(t)||t| + 3\|f\|_{L^1(\mathbb{R})} e^{-\frac{\text{dist}(S'' \setminus S')}{\Delta}}.$$ 

Proof. By the representation result in Proposition 5.6 and using the previous theorem, we have

$$\|f(H) - f(E_{S'}(B)HE_{S'}(B))E_{S'}(B)\| \leq \int_{\mathbb{R}} \|e^{itH} - e^{itE_{S'}(B)HE_{S'}(B)}E_{S'}(B)\| |\hat{f}(k)| dk \leq 2 \int_{|t| > \frac{\text{dist}(S'' \setminus S')}{e^2\Delta}} |\hat{f}(k)| dk + 3\|f\|_{L^1(\mathbb{R})} e^{-\frac{\text{dist}(S'' \setminus S')}{\Delta}}.$$ 

How we will use this is to form nonconsecutively orthogonal subspaces $Y_i$ where neighboring subspaces have significant overlap. Then for a vector $v$ in the span of these spaces, we can break it up into an orthogonal sum $\sum v_i$ where $v_i \in Y_i$ is well nested inside $Y_i$. Then we can guarantee that $f(H)v_i$ is approximately equal to $f(H_i)v_i$, where $H_i$ is $H$ restricted to $Y_i$. See the properties of the spaces $N_i$ below.

7 Davidson’s Reformulations of Lin’s Theorem

Davidson’s three equivalent reformulations of Lin’s theorem are:

(Q): For every $\epsilon > 0$ there is a $\delta > 0$ such that if $A, B$ are self-adjoint contractions with $\|A - B\| \leq \delta$, then there are commuting contractions $A', B'$ such that $\|A - A'\|, \|B - B'\| \leq \epsilon$.

(Q'): For every $\epsilon > 0$, there is an $L_0$ with the following property. If $J$ is a self-adjoint contraction that is block tridiagonal with respect to the following orthogonal subspaces $V_1, \ldots, V_{L_0}$, then there is a subspace $W$ such that $V_1 \subset W \subset \bigoplus_{i=1}^{L_0} V_i$ and $\|[J, P_W]\| \leq \epsilon$.

(Q''): For every $\epsilon > 0$, there is an $L_0$ with the following property. If $\mathcal{K}$ is a finite dimensional subspace of $L^2([0, 1])$ and $M_x$ is the multiplication operator by $x$ on $L^2([0, 1])$, then there is a subspace $W$ such that $\mathcal{K} \subset W \subset \text{span}_{0 \leq i \leq L_0 - 1} M_x^i(\mathcal{K})$ and $\|[J, P_W]\| \leq \epsilon$.

We now discuss the reduction of (Q) to (Q'). Although Section 3 of [5] does much of what we discuss in this section, we follow the notation and argument of [1], where Davidson’s reformulations are not explicitly mentioned. We start with $A, B$ self-adjoint with $\|A\|, \|B\| \leq 1$ and $\|(A, B)\| \leq \delta$. We will pick $\Delta = \delta^{10} << \delta$, but leave it as is for now (for simplicity and also for intuition).

We now proceed into Section III of [1], where we construct “the new basis”. Because $\sigma(A) \subset I = [-1, 1]$, we will cut up the interval $I$ into $n_{\text{cut}}$ (chosen later to equal $\lfloor 1/\Delta^{10} \rfloor$) for $0 < \gamma_1 < 1$ many disjoint intervals $I_i$ of the form $I_i = [-1 + i2/n_{\text{cut}}, -1 + (i + 1)2/n_{\text{cut}}]$ for $0 \leq i < n_{\text{cut}} - 1$ and $I_{n_{\text{cut}} - 1} = [1 - 2/n_{\text{cut}}, 1]$. Then we “pinch” $H$ by the projections $E_{I_i}(B)$ getting matrices $J_i$ acting on a spaces $B_{i+1} = \text{Ran}(E_{I_i}(B))$, $i = 0, \ldots, n_{\text{cut}} - 1$. 

"
Pictorially, we now focus only on $I_i$ and on that it has length $\frac{2}{n_{\text{cut}}}$. We will pick $n_{\text{cut}}$ later so that $\Delta = o(1/n_{\text{cut}})$ and hence we can partition $I_i$ into at least $\left\lfloor \frac{2/n_{\text{cut}}}{\Delta} - 1 \right\rfloor =: L$ many intervals $I_{i,j}$ of length at most $\Delta$ (only the first and last subinterval may have length less than $\Delta$). If $H$ has finite range $\Delta$, we obtain that $J_i = E_{i,L}(B)HE_{i,L}(B)$ is block tridiagonal with respect to the subspaces $V_{i,j}$ that the $E_{i,j}(B)$ project onto.

Naturally, given our choices above, we will get $L$ increasing like a negative power of $\Delta$, so either $J_i$ has many blocks or, because at least one block is empty because $V_{i,j} = 0$, we will obtain a nontrivial reducing subspace for $J_i$ such that the following lemma trivially holds.

Now, we state the main lemma for the argument in [1].

**Lemma 7.1.** Let $J$ be self-adjoint with $\|J\| \leq 1$ acting on $\mathcal{B}$ with $L$ orthogonal subspaces $\mathcal{V}_i$ with respect to which $J$ is block tridiagonal. Then there is a subspace $\mathcal{W}$ of $\mathcal{B}$ satisfying

1. For any $v \in \mathcal{V}_1, |P_{\mathcal{W}^\perp}(v)| \leq \epsilon_3|w|$.
2. For any $w \in \mathcal{W}, |P_{\mathcal{W}^\perp}(Jw)| \leq \epsilon_4|w|$.
3. For any $w \in \mathcal{W}, |P_{\mathcal{L}}(w)| \leq \epsilon_5|w|$, where $i = 3, 4, \epsilon_i = \frac{|E_i(\frac{1}{L})|}{L}$ where $E_i(t)$ grows slower than any (positive) power (of $t$) and $\epsilon_5$ decays faster than any power of $L$.

The “construction” in the proof only works if $L$ is large, which is masked by the undefined nature of the $E_i$ in the lemma above, because we can imagine defining $E_i(1/L)$ to be large for all $L$ small. How large $L$ needs to be is undetermined because of the non-constructive step in the proof discussed in Section 11.

**Remark 7.2.** Intuitively, Item 1 above means that $\mathcal{V}_1$ is almost contained in $\mathcal{W}$, Item 2 means that $J$ is almost invariant under $\mathcal{W}$, and Item 3 means that $\mathcal{W}$ is almost orthogonal to $\mathcal{V}_L$.

**Remark 7.3.** Note that Item 3 above is formulated differently in [1], but they are equivalent because they each say that for all $w \in \mathcal{W}, v \in \mathcal{V}_L, |(w, v)| \leq \epsilon_5|w||v|$. The form stated above is what is proved in [1], whose proof we follow.

Similarly we have a dual statement for Lemma 2:

1. For any $w \in \mathcal{W}^\perp, |P_{\mathcal{V}_1}(w)| \leq \epsilon_3|w|$.
2. For any $w \in \mathcal{W}^\perp, |P_{\mathcal{V}}(Jw)| \leq \epsilon_4|w|$.
3. For any $v \in \mathcal{V}_L, |P_{\mathcal{V}}(v)| \leq \epsilon_5|v|$.

where the second statement uses that $J$ is self-adjoint.

Although the following result is what we will use in the proof of Lin’s theorem (because it more closely follows Davidson’s and Szarek’s results), Lemma 7.1 is called the main lemma, because most of this paper is dedicated to proving it.

**Lemma 7.4.** Let $J$ be self-adjoint with $\|J\| \leq 1$ acting on $\mathcal{B}$ with $L$ orthogonal subspaces $\mathcal{V}_i$ with respect to which $J$ is block tridiagonal. Then there is a subspace $\mathcal{W}$ of $\mathcal{B}$ satisfying $\mathcal{V}_1 \leq \mathcal{W} \perp \mathcal{V}_L$ and

$$\|P_{\mathcal{W}^\perp}JP_{\mathcal{W}}\| \leq \epsilon_4 + 10\max(\epsilon_3, \epsilon_5) =: \epsilon_2,$$

where $\epsilon_3, \epsilon_4, \epsilon_5$ are as in Lemma 7.1.
Figure 1: The formation of the new basis from the old. Note that the interval illustrates the
197 DAVIDSON’S REFORMULATIONS OF LIN’S THEOREM

Remark 7.5. The proof of Theorem 3.2 of [5] (the equivalence of this lemma and Lin’s
theorem) shows that we can pick $\epsilon_2 = 1/L^{1/2}$ by simply applying Lin’s theorem.

This shows that we obtain the “local joint near-diagonalization” described in the
introduction.

Now, using Lemma 7.4 we obtain Lin’s theorem as detailed in [1]. Consider “the
new basis” of subspaces for $0 \leq j \leq n_{cut}$,

$\tilde{B}_j : \mathcal{W}_1, \mathcal{W}_1 \perp \mathcal{W}_2, \ldots, \mathcal{W}_{i} \perp \mathcal{W}_{i+1}, \ldots, \mathcal{W}_{n_{cut}-1} \perp \mathcal{W}_{n_{cut}}, \mathcal{W}_{n_{cut}}$.

See Figure 1 below. Note the abuse of notation that we will use for the rest of the

section: $\mathcal{W}_i^\perp = B_i \ominus \mathcal{W}_i$ and $B_i^\perp = B \ominus B_i$.

For simplicity, set $\mathcal{W}_0 = \mathcal{W}_0^\perp = B_0 = 0$ and $I_{n_{cut}} = \{1\}$ so that for $0 \leq i \leq n_{cut}$,

$\tilde{B}_i = \mathcal{W}_i^\perp \ominus \mathcal{W}_{i+1} \subset B_i \ominus B_{i+1}$. Now, let $B'$ be the block identity matrix which equals
the left endpoint of $I_i$ multiplied by the identity on $\tilde{B}_i \subset B_i \ominus B_{i+1}$, $0 \leq i \leq n_{cut} - 1$.
Because $B$ has eigenvalues in $I_I$ on $B_{I+1}$, we see that $\|B - B'\| \leq \frac{2}{n_{cut}}$.

We now show that the $\tilde{B}_i$ are almost invariant subspaces for $H$, where we let $H'$
be the pinching of $H$ along $\tilde{B}_i$ so then $\|H - H'\| \leq 2\epsilon_2$. This amounts to showing that
applying $H$ to any vector in $\mathcal{W}_i$ remains in $\mathcal{W}_{i-1} \perp \mathcal{W}_i$ except that very small amounts
are permitted to “leak out”. A similar statement would hold for $\mathcal{W}_{i-1}^\perp$. The idea is
that applying $H$ to $\mathcal{W}_i \subset \tilde{B}_i$ can only make $\mathcal{W}_i$ leak out into $\mathcal{W}_{i+1}^\perp$ and $\mathcal{W}_{i+1}$ through
$\mathcal{V}_L^i$ and $\mathcal{V}_L^i$, respectively. In details, let $B_i$ be written as the orthogonal direct sum of
$\mathcal{V}_1^i, \ldots, \mathcal{V}_L^i$ as in the statement of Lemma 2.

Recall that by Item 2, $\|P_{\mathcal{W}_i} H P_{\mathcal{W}_i}\| \leq \epsilon_2$. By the block tridiagonality of $H$,$P_{\tilde{B}_i} H P_{\mathcal{W}_i} = P_{\tilde{B}_i} H (P_{\mathcal{V}_i} + P_{\mathcal{V}_L}) P_{\mathcal{W}_i}$. Now, $P_{\mathcal{V}_L} P_{\mathcal{W}_i} = 0$ and $P_{\tilde{B}_i} H P_{\mathcal{V}_i} P_{\mathcal{W}_i} = P_{\tilde{B}_i} H P_{\mathcal{V}_i}$.
maps into $V_{L}^{i-1} \subset W_{L}^{i-1}$ so $P_{B_{i}}H P_{W_{i}} = P_{V_{L}^{i-1}}H P_{W_{i}} = P_{W_{L}^{i-1}}H P_{W_{i}}$. Hence,

$$
\|(1 - P_{B_{i}})H P_{B_{i}}P_{W_{i}}\| = \|H P_{W_{i}} - P_{W_{i}}H P_{W_{i}} - P_{W_{L}^{i-1}}H P_{W_{i}}\|
\leq \|P_{B_{i}}H P_{W_{i}} - P_{W_{i}}H P_{W_{i}}\| + \|P_{B_{i}}H P_{W_{i}} - P_{W_{L}^{i-1}}H P_{W_{i}}\| \leq \epsilon_{2} + 0.
$$

Likewise, by taking adjoints, $\|P_{W_{i}}H P_{W_{i}}\| \leq \epsilon_{2}$. By the block triangularity of $H$, $P_{B_{i}}H P_{W_{i}} = P_{B_{i}}H (P_{V_{1}} + P_{V_{L}}) P_{W_{i}}$. Now, $P_{V_{1}}P_{W_{i}} = 0$ and $P_{B_{i}}H P_{V_{L}} P_{W_{i}} = P_{B_{i}}H P_{V_{L}}$ maps into $V_{i+1} \subset W_{i+1}$, so $P_{B_{i}}H P_{W_{i}} = P_{V_{i+1}}H P_{W_{i}} = P_{W_{i+1}}H P_{W_{i}}$. Hence,

$$
\|(1 - P_{B_{i}})H P_{B_{i}}P_{W_{i}}\| = \|H P_{W_{i}} - P_{W_{i+1}}H P_{W_{i}} - P_{W_{L}^{i}}H P_{W_{i}}\|
\leq \|P_{B_{i}}H P_{W_{i}} - P_{W_{i+1}}H P_{W_{i}}\| + \|P_{B_{i}}H P_{W_{i}} - P_{W_{L}^{i}}H P_{W_{i}}\| \leq 0 + \epsilon_{2}.
$$

So, it follows that

$$
\|P_{B_{i}}H P_{B_{i}}\| \leq 2\epsilon_{2}.
$$

Set $H' = \sum_{j} P_{B_{j}}H P_{B_{j}}$. Because the spaces $B_{i}$ are orthogonal, we see that $\|H - H'\| \leq 2\epsilon_{2}$. By construction $B'$ and $H'$ commute and we conclude.

### 8 General Approach to the construction of $\mathcal{W}$.

In this section we define subspaces $\tilde{X}_{i}$ and $\tilde{X}'_{i}$ to describe the motivating ideas used in [17]. We then define smooth cut off functions $F_{\omega_{0}}^{\omega_{W}}(t)$ and use them to define the spaces $X_{i}$, which will be put together in some sense to define $\mathcal{W}$ following [1].

Because we need to essentially recover $V_{1}$ as a subspace of $\mathcal{W}$ by Item 1, we might consider finding subspaces $\tilde{X}_{i}$ such that $V_{1}$ is approximately a subset of the sum of the $\tilde{X}_{i}$. For example, consider breaking up $[-1, 1]$ into $n_{\text{win}}$ many disjoint intervals $I_{i}$ similarly to how we did before then we could consider $\tilde{X}_{i} = [\chi_{I_{i}}(J)](V_{1})$. This space recovers $V_{1}$ because any element $v$ in $V_{1}$ can be written $v = \sum_{i} x_{i}$, where $x_{i} = \chi_{I_{i}}(J)v$.

These spaces are almost invariant: Let $\omega(i)$ be the midpoint of $I_{i}$ and $x \in \tilde{X}_{i}$. We have that

$$
|(J - \omega(i))x| = |(J - \omega(i))\chi_{I_{i}}(J)x| \leq \|(J - \omega(i))\chi_{I_{i}}(J)\| |x| \leq \frac{|I_{i}|}{2} |x|.
$$

This shows that $Jx$ is almost in span($x$) $\subset \tilde{X}_{i}$, where the error depends on the norm of $x$. Also, these spaces $\tilde{X}_{i}$ are also orthogonal, so we obtain that if $v = \sum_{i} x_{i}$ then $|v|^{2} = \sum_{i} |x_{i}|^{2}$.

Putting all of this together, if we set $\tilde{\mathcal{W}} = \bigoplus_{i} \tilde{X}_{i}$, we have Item 1 because it contains $V_{1}$. It satisfies Item 2 because if $w \in \tilde{\mathcal{W}}$ then we can write it as $w = \sum_{i} x_{i}$, $x_{i} \in \tilde{X}_{i}$. The $Jx_{i}$ are orthogonal because the $I_{i}$ are disjoint. So,

$$
|Jw - \sum_{i} \omega(i)x_{i}|^{2} = |\sum_{i}(J - \omega(i))x_{i}|^{2} = \sum_{i}|(J - \omega(i))x_{i}|^{2}
\leq \sum_{i} \left(\frac{|I_{i}|}{2} \right)^{2} |x_{i}|^{2} = \left(\frac{|I_{i}|}{2} \right)^{2} |w|^{2}.
$$
so because \( x_i \in \tilde{W} \),

\[
|P_{\tilde{W}^\perp} Jw| = |P_{\tilde{W}^\perp} \left(Jw - \sum_i \omega(i)x_i\right)| \leq \frac{|I|}{2}|w|.
\] (10)

This general set-up is common ground for Szarek’s and Hastings’ arguments. They differ in how to modify this core argument to get a result concerning Item 3. We proceed with a modification used in both constructions.

We want to show that Item 3 holds as well for \( \tilde{W} \), but it might not unless we address an issue. Recall that \( \tilde{X}_i = \{\chi_I(J)v : v \in V_i\} \). Let \( S_1 \) be a matrix whose columns are a fixed orthonormal basis of \( V_1 \), where \( d = \dim V_1 \). Then \( S_1 \) is an isometric isomorphism between \( \mathbb{C}^d \) and \( V_1 \) and \( \operatorname{Ran}(\chi_I(J)S_1) = \tilde{X}_1 \).

We might hope that we have exponential decay of the columns of \( S_1 \) (which would provide the result that we want) in a way similar to Proposition 2 or as mentioned in Remark 4.3. This might not happen. To see this, \( \square \) gives an example of an \( n \times n \) matrix of the form:

\[
\begin{pmatrix}
0 & 1/4 & 0 \\
1/4 & 0 & 1/4 & 0 \\
0 & 1/4 & 0 & 1/4 & \ddots \\
& & & & & 0 \\
& & & & & 0 \\
& & & & & 0 \\
& & & & & 0 \\
& & & & & 0 \\
& & & & & 0 \\
& & & & & 0
\end{pmatrix}
\]

Here, we can take \( V_1 \) to be the first basis vector. For \( n \) large, its spectrum is distributed finely in \([-1/2, 1/2] \) with a single (with multiplicity one) eigenvalue at around 5/8.

For example, MATLAB calculations gives that if \( S_1 \) is just the first standard basis vector then \( \chi_{[5/8-1/100,5/8+1/100]}(J)S_1 \) is, for \( n = 10 \),

\[
10^{-3} \times (0.0016, 0.0040, 0.0084, 0.0171, 0.0343, 0.0686, 0.1373, 0.2747, 0.5493, 1.0987)^T,
\]

and for \( n = 50 \)

\[
10^{-15} \times (0.0000, \ldots, 0.0000, 0.0001, 0.0002, 0.0005, 0.001, 0.002, 0.004, 0.008, 0.016, 0.031, 0.062, 0.125, 0.251, 0.502, 1.004)^T.
\]

The pattern is that the norm of this vector is very small, but does not have small projection onto \( V_L \). One way of addressing this phenomenon is to disallow vectors that have too small norm, by removing columns of \( \chi_I(J)S \) that are too small. However, that does not exclude the case that all of its columns are large, but a linear combination of them are of this problematic type. For example, we might have chosen a different basis for \( V_1 \). If one modifies the operators \( \chi_I(J)S \) so that we still maintain Item 1 and Item 2 and removing the issue discussed above, we might be closer to showing Item 3.

The approach to this problem is to write \( \tilde{\tau}_i = \chi_I(J)S_1 \) and remove its singular values that are too small by letting \( \tilde{Z}_i \) project onto the eigenspace of \( \tilde{\tau}_i^2 \tilde{\tau}_i \) for eigenvalues less than some \( \lambda_{\min} \). Then define new spaces \( \tilde{X}_i' \) to be the range of \( \tilde{\tau}_i(1 - \tilde{Z}_i) \) and let \( W' = \bigoplus_i \tilde{X}_i' \), an orthogonal direct sum.
We return to the general case. This definition of $\tilde{X}_i^j$ gives that for any $x \in \tilde{X}_i^j$, we have some $x \in \mathbb{C}^d$ such that $\tilde{\tau}_i(1 - \tilde{Z}_i)x = x$ and we can choose $x$ to be in the range of $1 - \tilde{Z}_i$ so that $\tilde{\tau}_i x = x$. We call $x$ the representative of $x$ (by $\tilde{\tau}_i$). Writing $x = \sum_a x^a$ as the orthogonal sum of eigenvectors $x^a$ of $\tilde{\tau}_i^* \tilde{\tau}_i$ with eigenvalues $\lambda_a \geq \lambda_{\min}$ we obtain

$$|\tilde{\tau}_i x|^2 = \sum_{a,b} (\tilde{\tau}_i x^a, \tilde{\tau}_i x^b) = \sum_a (\tilde{\tau}_i^* \tilde{\tau}_i x^a, x^a) \geq \sum_a \lambda_{\min} |x^a|^2,$$

so

$$|\tilde{\tau}_i x| \geq \lambda_{\min}^{1/2} |x|.$$

This guarantees that we do not have exponentially small “problematic” vectors if we make $\lambda_{\min}$ go to zero only like a power of $L$. This adjustment seemingly does not offer a way to prove Item 3, however, if it were true.

Item 1 holds for $W'$ because for any $v \in V_1$, we can write it as the orthogonal direct sum $\sum_i x_i$, where $x_i = \chi_{I_i}(J)v = \tilde{\tau}_i x$ and $v = Sx$. Then we get that

$$|\tilde{\tau}_i x - \tilde{\tau}_i(1 - \tilde{Z}_i)x|^2 = |\tilde{\tau}_i \tilde{Z}_i x|^2 \leq \lambda_{\min} |x|^2 = \lambda_{\min} |v|^2.$$

Because the ranges of the $\tilde{\tau}_i$ are orthogonal, we get that

$$|P_{W'} v| \leq \left| \sum_i \tilde{\tau}_i x - \sum_i \tilde{\tau}_i(1 - \tilde{Z}_i)x \right| = \left| \sum_i \tilde{\tau}_i \tilde{Z}_i x \right| = \left( \sum_{i=0}^{n_{\min} - 1} |\tilde{\tau}_i \tilde{Z}_i x|^2 \right)^{1/2} \leq \left( n_{\min} \lambda_{\min} \right)^{1/2} |v|.$$

We see that Item 2 holds as follows. Let $\omega(i)$ be the midpoint of the interval $I_i$. For $w = \sum_i x_i$ with $x_i \in \tilde{X}_i^j$ so $\chi_{I_i}(J)x_i = x_i$, we have

$$|(J - \omega(i)) x_i | \leq \frac{|I_i|}{2} |x_i|.$$

Recall that $x_i \in \tilde{X}_i^j \subset \tilde{X}_i$ and $Jx_i \in \tilde{X}_i$. So, because the $\tilde{X}_i$ are orthogonal we obtain

$$|P_{W'} Jw|^2 = |P_{W'} \sum_i (J - \omega(i)) x_i|^2 \leq \sum_i |(J - \omega(i)) x_i|^2 = \sum_i |(J - \omega(i)) x_i|^2 \leq \sum_i \left( \frac{|I_i|}{2} |x_i| \right)^2 \leq \left( \frac{|I_i|}{2} |w| \right)^2$$

so $|P_{W'} Jw| \leq \frac{|I_i|}{2} |w|$.

An important property that we used in approximating $Jw$ was that $Jw = \sum_i Jx_i$ and not only is $\sum x_i = w$ an orthogonal decomposition but the $(J - \omega(i)) x_i$ are also orthogonal because the $\tau_i$ are. Later when we define $X_i$ to be a subspace of the range of $J_{\omega(i)}^0$ (where $J_{\omega(i)}^0$ are smooth overlapping cut-off functions), we see that $J(X_i^j)$ are nonconsecutively orthogonal. One could see that the above argument would work, except that we seemingly do not have

$$\sum_i |x_i|^2 \leq \text{Const.} |w|^2. \quad (11)$$
These are some of the obstructions that Hastings deals with using the $F_{\omega(i)}^{0,k}$. The subspaces that we obtain will no longer be orthogonal (which causes its own issues), but in Lemma 15.2 we obtain the estimate
\[ \sum_i |x_i|^2 \leq \text{Const}.l_b|w|^2 \]
where $l_b$ grows like a power of $L$ and along with the decay given in Corollary 6.7 we can obtain the third property. The smoothness of $F_{\omega(i)}^{0,k}$ ensures the sufficient decay of its Fourier transform and hence that of Corollary 6.7. As stated in [1], “The smoothness will be essential to ensure that the vectors ... have most of their amplitude in the first blocks rather than the last blocks”.

### 9 Constructions in Certain Cases

One comment is that Lin’s theorem is essentially the finite dimensional case if we assume that $A$ always has at most $m$ (distinct) eigenvalues. If there are finitely many eigenvalues, this implies dimensional dependent result of $\delta = \frac{1}{2m^2}\epsilon^2$, in the case of the following cheap result.

**Proposition 9.1.** Let $A, B \in M_n(\mathbb{C})$, where $A$ is self-adjoint with at most $m$ distinct eigenvalues. Then there are commuting $A', B'$ such that $A'$ is self-adjoint and
\[ \|A - A'\|, \|B - B'\| \leq \frac{m}{\sqrt{2}}\|A, B\|^{1/2}. \]

**Proof.** We then partition the spectrum of $A$ into intervals $I_i$ of length at most $c_m \delta^{1/2}$, where $\delta := \|A, B\|$. We choose $c_m$ later so as to make our bound for $\|A - A'\|$ equivalent in size to that of $\|B - B'\|$. We write $A$ as block multiples of the identity $(a_i I_i)$, with the $a_i$ ordered increasingly, and $B$ as a block matrix $(B_{ij})$, so then $[A, B]_{rs} = (a_r - a_s)B_{rs}$.

We cannot avoid that some off-diagonal terms of $(B_{ij})$ might not be small where the eigenvalues of $A$ are close because of behavior depicted in Example 5.5. A way of dealing with this is to use the trick in Hastings’ proof of the tridiagonal case and merge together intervals that are too close. This will be doable, because our estimate will depend on the number of eigenvalues.

Doing this, take our partition of half-open intervals $I_i$ of length $c_n \delta^{1/2}$ and discard the intervals that do not intersect the spectrum of $A$ to obtain half-open intervals $I_j$. Merge together neighboring intervals to obtain merged intervals $I_k$. Because there were at most $m$ eigenvalues, these merged intervals have length at most $mc_m \delta^{1/2}$ and there is a gap of at least $c_m \delta^{1/2}$ (the length of one $I_i$). We let $A'$ be the block matrix formed by merging the diagonal entries of $A$ that correspond to the merged intervals and replacing the entries with the midpoint of the merged interval. Then $\|A - A'\| \leq \frac{1}{2}mc_m \delta^{1/2}$. We let $B'$ be the matrix gotten by discarding the off-diagonal blocks $B_{rs}$ where $a_r$ and $a_s$ are from different merged blocks. Because $|a_r - a_s||B_{rs}| \leq ||[A, B]|| = \delta$ and for $a_r, a_s$ from different merged blocks, $|a_r - a_s| \geq c_m \delta^{1/2}$ so we obtain for these $r, s$, $\|B_{rs}\| \leq \frac{1}{c_m} \delta^{1/2}$.

Because matrices in $M_n(\mathbb{C})$ are being expressed as square block matrices with at most $m$ rows, $\|B - B'\| \leq \frac{m}{c_m} \delta^{1/2}$. Setting $c_m = \sqrt{2}$, we obtain
\[ \|A - A'\|, \|B - B'\| \leq \frac{m}{\sqrt{2}} \delta^{1/2}. \]
Remark 9.2. It should be said that this result is (intentionally) very similar to Pearcy and Shield’s result in [9]. In fact, that result still holds under our weaker condition that there are \( m \) eigenvalues of \( A \) which gives \( \| A - A' \|, \| B - B' \| \leq \left( \frac{m - 1}{2} \| [A, B] \| \right)^{1/2} \) by their method of more carefully picking “merged intervals” and estimating the norm of \( \| B - B' \| \).

Remark 9.3. Davidson’s counter-example of \((n^2 + 1) \times (n^2 + 1)\) self-adjoint \( A \) and almost normal \( B \) that satisfy \( \|[A, B]\| = 1/n^2 \) and have no nearby commuting matrices shows that [6]’s result is strict.

The choice of the groupings of eigenvalues of \( A \) does not involve \( B \) in any way, besides the use of \( \|[A, B]\| \). A difficulty in solving Lin’s theorem involves the fact that we need to use \( B \) when constructing \( A' \) (and vice-versa). This is a consequence of the following result.

Proposition 9.4. Suppose that \( A_k, B_k \in M_{n_k}(\mathbb{C}) \) are self-adjoint contractions. If \( \|[A_k, B_k]\| \to 0 \), by Lin’s theorem there are self-adjoint, commuting \( A_k', B_k' \) so that \( \| A_k - A_k' \|, \| B_k - B_k' \| \to 0 \). Let \( C_k \) be the number of eigenvalues of \( A_k' \).

If the choice of \( A_k' \) can be made independent of \( B_k \), then \( \sup_k C_k < \infty \).

Proof. We prove this by contradiction. Suppose that \( \sup_k C_k = \infty \). We choose a subsequence and relabel so that \( C_k \geq 2k \).

By the Pigeonhole principle, there are two distinct eigenvalues \( \lambda_k^1, \lambda_k^2 \) of \( A_k' \) with eigenvectors \( v_k^1, v_k^2 \), respectively, so that \( |\lambda_k^1 - \lambda_k^2| \leq 1/k \). If we define \( B_k \) to be the linear operator satisfying \( B_k v_k^1 = v_k^2, B_k v_k^2 = v_k^1 \), and is identically zero on the orthogonal complement of \( \text{span}(v_k^1, v_k^2) \). Then \( \|[A_k', B_k]\| \leq 1/k \to 0 \).

If \( B_k' \) is any matrix that commutes with \( A_k' \) then the eigenspaces of \( A_k' \) are invariant under \( B_k' \). This means that when writing \( B_k' \) as a block diagonal matrix with respect to the eigenspaces of \( A_k' \), it must be diagonal. This shows that \( \|B_k - B_k'\| \geq 1 \). \( \square \)

Remark 9.5. Compare to Example 5.5.

The following result illustrates what seems to be a rather strict requirement on the size of the commutators because it destroys the invariants for three almost commuting Hermitians. Compare to Remark 9.3. Unlike [9], this result is not constructive.

Corollary 9.6. Let \( A_k, B_k, C_k \in M_{n_k}(\mathbb{C}) \) be self-adjoint contractions with \( A_k \) having \( m_k \) (distinct) eigenvalues. If \( \|[A_k, B_k]\|, \|[A_k, C_k]\| = o(1/m_k) \) and \( \|[B_k, C_k]\| = o(1) \) then there are self-adjoint commuting \( A_k', B_k', C_k' \) such that

\[
\|A_k - A_k'\|, \|B_k - B_k'\|, \|C_k - C_k'\| = o(1).
\]

Proof. Note that \( o(1) \) only depends on \( k \).

Applying [6]’s construction to \( A_k \), we get disjoint intervals \( I_{k1}, \ldots, I_{kn} \) that cover \( \sigma(A_k) \). For the projections \( P_{i_k}^k = E_{i_k}(A_k) \), we can modify the spectrum of \( A_k' := P_{i_k}^k A_k P_{i_k}^k \) to get \( (A_k')^i \) multiples of identity matrices which form the diagonal entries of the block diagonal matrix \( A_k' \). This construction satisfies the following properties:

\[
\|A_k - A_k'\| \leq ((m_k - 1) \max(\|[A_k, B_k]\|, \|[A_k, C_k]\|))/2)^{1/2} =: \epsilon_k = o(1).
\]
and for $B''_i = \sum_j P^k_i B_k P^k_i$, $C''_i = \sum_j P^k_i C_k P^k_i$ we have $\|B_k - B''_i\|, \|C_k - C''_i\| \leq \varepsilon_k = o(1)$. Then

$$\|\langle B''_i, C''_i \rangle\| \leq \|\langle B_k, C_k \rangle\| + 2\|B_k - B''_i\| + 2\|C_k - C''_i\| = o(1).$$

By definition,

$$\|\langle B''_i, C''_i \rangle, P^k_i\| = \|P^k_i \langle B''_i, C''_i \rangle P^k_i\| = o(1)$$

and so we can apply Lin’s theorem to these block matrices, so there are commuting matrices $B'_k, C'_k$ of the same block size as $P^k_i B_k P^k_i$, $P^k_i C_k P^k_i$ such that $\|B'_k - P^k_i B'_k P^k_i\|, \|C'_k - P^k_i C'_k P^k_i\| = o(1)$. Since $P^k_i A'_k P^k_i$ is a multiple of the identity, we see that forming $B'_k = \bigoplus_i B'_i C'_k$, $C'_k = \bigoplus_i C'_i$ gives us commuting self-adjoint matrices $A'_k, B'_k, C'_k$ satisfying the statement of the lemma.

Now, we proceed to look at cases where we can get explicit bounds in Lemma 7.4. As a simplified case of Hastings’s general proof, Hastings proves a Lemma 7.1-type result for tridiagonal matrices. The idea is that although some spaces $X_i$ are not necessarily orthogonal, one can group together the $X_i$ so that there is control on how non-orthogonal these spaces are (in the sense of $C = 1/3$) and then instead of using the Pythagorean theorem we would use (3). What makes this possible is that the $X_i$ are one dimensional so either we have (2) for all vectors $v_i, v_{i+1}$ in $X_i, X_{i+1}$, respectively, or we have the opposite inequality. Besides this, the set up for the proof of the general case of Hastings is rather similar. However, because Szarek in [17] states a result that solves the problem for 2m-banded matrices that is not explicitly put in those terms, we will sketch this earlier method that has much similarity to the general approach taken in [1].

A relevant result for tridiagonal matrices is that along the way of proving his result, Szarek states a proposition that gives Lin’s theorem when one block size of the block tridiagonal matrix is controlled. This gives him a main result of [17]:

**Theorem 9.7.** There is a universal constant $c > 0$ such that if self-adjoint contractions $A, B \in M_n(\mathbb{C})$ have rank at most $m$, then there are self-adjoint $A', B' \in M_n(\mathbb{C})$ such that

$$\|A - A'\|, \|B - B'\| \leq c(m^{1/2}\|A, B\|)^{2/13}.$$

As Szarek mentions, the key result was improving Pearcy and Shield’s result so that one gets a factor of $m^{1/2}\|A, B\|$ instead of $m\|A, B\|$.

The proposition in [17] states that one gets a Lemma 7.4-type result when one has control of $m = \min_j \dim V_j$. To see this, we follow (and modify) the construction originally from [5]. This is done by reducing to the case where all blocks have size $m = \min(m_-, m_+)$, if there is some $i$ where $J(V_i)$ has rank $m_-$ projected onto $V_{i-1}$ and rank $m_+$ projected onto $V_{i+1}$. Note that because $J$ is self-adjoint, rank $P_{V_i} J P_{V_i}$ = rank $P_{V_i} J P_{V_{i+1}}$, so we can just assume that for some $i < L$, $m = \dim P_{V_i} J(V_i)$.

We will prove this result when $i \leq \lfloor L/2 \rfloor$. If $i > \lceil L/2 \rceil$, then we note that $m = \dim P_{V_i} J(V_{i+1})$ and consider that $J$ is block tridiagonal with respect to the (oppositely indexed) blocks $V_{1}^R, \ldots, V_{L}^R : V_L, \ldots, V_1$ so

$$m = \text{rank } P_{V_i} J P_{V_{i+1}} = \text{rank } P_{V_{L-i+1}}^R J P_{V_{L-i}}^R.$$

Because $L - i \leq L - \lfloor L/2 \rfloor \leq L/2$, we conclude from the case that we are about to do. So, we can just assume that $i \leq \lfloor L/2 \rfloor$. 


Define $\mathcal{M}_{-1} = 0, \mathcal{M}_0 = \bigoplus_{k=1}^i \mathcal{V}_k, \mathcal{M}_{k+1} = \text{span}(\mathcal{M}_k, J(\mathcal{M}_k))$ and $\mathcal{H}_{k+1} = \mathcal{M}_{k+1} \ominus \mathcal{M}_i$. We get a finite chain of subspaces $\mathcal{H}_0, \ldots, \mathcal{H}_{n_+}$, where $n_+ \geq 1$ is the last number $k$ such that $\mathcal{H}_k \neq 0$. Then the spaces $\mathcal{H}_k$ are orthogonal, $\bigoplus_{k=0}^{n_+} \mathcal{H}_k$ is an invariant subspace for $J$, and because $J$ is self-adjoint these subspaces form a block tridiagonal structure for $J$ as $J(\mathcal{H}_k) \perp \mathcal{H}_{k+2}$. Note that for $k \geq 1$, $\mathcal{M}_k = \text{span}(\mathcal{M}_0, J(\mathcal{V}_i), \ldots, J^k(\mathcal{V}_i))$ due to the tridiagonal nature of $J$ with respect to the $\mathcal{V}_j$. Because $J(\mathcal{V}_i)$ has dimension $m$, this implies that $\mathcal{H}_k$ has dimension at most $m$ for $k \geq 1$.

Note that $\mathcal{H}_k \subset \mathcal{M}_k \subset \bigoplus_{j=1}^{i+k} \mathcal{V}_j$, so $\mathcal{H}_k \perp \mathcal{V}_L$ for $k < L - i$. Consequently, if $i + n_+ < L$, then $\mathcal{M}_{n_+}$ is orthogonal to $\mathcal{V}_L$, in which case the proof of Lemma 7.4 is trivial: we pick $W = \bigoplus_{k=0}^{H} \mathcal{H}_k$, which is orthogonal to $\mathcal{V}_L$ and reduces $J$. So, we can assume that $n_+ \geq L - i \geq (L - 1)/2$.

Let $J_{i,+}$ be the projection of $J$ onto $\bigoplus_{k\geq i} \mathcal{H}_k$. Then $J_{i,+}$ is tridiagonal with respect to the blocks $\mathcal{H}_k$, $k \geq 1$ and each of the blocks has dimension at most $m$. If we can prove the lemma when all the blocks have dimension at most $m$, we can find a subspace $W \subset \bigoplus_{k=1}^{n_k} \mathcal{H}_k$ such that $\mathcal{H}_1 \leq W \perp \mathcal{H}_{L-i}$ and $\|J_{i,+}^k P_{\mathcal{H}_k \cap W} \| \leq \epsilon_1((L-1)/2)$. Then we see that setting $W = \mathcal{H}_0 \ominus W$ gives the result, because $J$ maps $\mathcal{H}_0$ into $\mathcal{H}_1 \perp \mathcal{H}_1 \subset W$ so $P_{W \perp J} P_W = P_{\bigoplus_{k=0}^{i-1} \mathcal{H}_k \cap W} J_{i,+} P_W$.

The proof of “Proposition” from [17] continues now as before. The point of this comment is that not only do we get good bounds for a matrix that contains a block like this:

$$
\begin{pmatrix}
  * & 1 & * & * & * \\
  1 & * & * & * & * \\
  * & * & * & * & * \\
  * & * & * & * & * \\
  * & * & * & * & 1 \\
\end{pmatrix}
$$

because there is a block of size one, but also for a matrix that contains a block like this:

$$
\begin{pmatrix}
  * & * & * & * & * \\
  * & * & * & * & * \\
  * & * & * & * & * \\
  * & * & * & * & * \\
  * & * & * & * & 1 \\
\end{pmatrix}
$$

We summarize this discussion in a modified version of Proposition from [17]. Note that this result is constructive.

**Proposition 9.8.** Let $J$ be self-adjoint with $\|J\| \leq 1$ acting on $\mathcal{B}$ with $L$ orthogonal subspaces $\mathcal{V}_i$ with respect to which $J$ is block tridiagonal. If for some $i < L$ we set $m = \text{rank } P_{\mathcal{V}_L} J P_{\mathcal{V}_L}$, then there is a subspace $W$ such that $\mathcal{V}_1 \subset W \perp \mathcal{V}_L$ and

$$
\|P_{W \perp J} P_W\| \leq \epsilon_1,
$$

where $M$ is a constant from the approximation of polynomials in Lemma 2 of [17] and $\epsilon_1 = 83.4 \left( \frac{mM}{L^2 - 2} \right)^{1/9}$ for $L_0 \geq \sqrt{2}M \max(m, 44^{9/5}m^{-4/5}) + 2$.

**Remark 9.9.** If $m = 1$, we can apply Lemma 6 of [1] to obtain $\epsilon_1 = E(1/L)^{1/2}$. 

26
We sketch the construction as follows. We start with a short lemma whose proof is sketched in [17] without the below stated condition on the relative size between \(\kappa\) and \(\eta\). This ensures that there are at least \(\kappa/4\eta\) many intervals of length \(\eta\) in any interval of length \(\kappa/2\) from which an interval of length \(\eta\) has been removed.

**Lemma 9.10.** Let \(\mu\) be a finite positive measure on \([0, 1]\) and let \(\kappa > 8\eta > 0\). Then there are disjoint intervals \(I_1, \ldots, I_r\) such that

1. \(r \leq 2\kappa^{-1}\)
2. \(\text{diam} I_j \leq \kappa\)
3. \(\text{dist}(I_i, I_j) \geq \eta, i \neq j\)
4. \(\mu \left([0, 1] \setminus \bigcup_j I_j \right) \leq \frac{4\eta}{\kappa} \mu([0, 1])\)

The argument in [17] is framed in terms of Davidson’s third formulation of Lin’s theorem, however the proof readily translates into the second formulation. This matrix perspective is more in line with [1] and gives a definition of \(\Phi\) below that is much more natural (although the measures and some calculations are more natural in [17]’s perspective is more in line with [1] and gives a definition of \(\Phi\) below that is much more natural (although the measures and some calculations are more natural in [17]’s notation). We use the matrix formulation.

Using the technique described above, we can assume that \(\mathcal{V}_1\) has dimension \(m\). For \(\epsilon > 0\), we want to construct a projection \(P\) such that \(P\mathcal{V}_1 \leq P \leq P_{\Phi^{\epsilon m^{-1}}\mathcal{V}_1}\) and \(\| [J, P] \| < \epsilon\).

Because [17]’s proof is framed in terms of the multiplication operator \(M_x\) on the Hilbert space \(L^2([0, 1])\), we first replace \(J\) with \((J + I)/2\) so that \(J\) is positive with norm at most one. Also, let \(\lambda_1 \leq \lambda_2 \leq \ldots\) be a list of the eigenvalues of \(J\), repeated with multiplicity with respective orthonormal vectors \(e_1, e_2, \ldots\). Let \(Q\) project onto \(\mathcal{V}_1\) and let \(\Phi\) be the vector \(\sum_k (\langle e_k | Q | e_k \rangle)^{1/2} | e_k \rangle\) and \(\Phi_k = \langle e_k | \Phi = \langle e_k | Q | e_k \rangle^{1/2}\), where we have used bra-ket notation. If \(f \in \mathcal{V}_1\) has norm one, the components of \(f\) in the basis \(e_1, e_2, \ldots\) can be bounded as follows:

\[
|f_k|^2 = |\langle e_k | f \rangle|^2 = |\langle e_k | f \rangle| |f| |e_k\rangle| \leq |\langle e_k | Q | e_k \rangle| = \Phi_k^2
\]

and also

\[
|\Phi|^2 = \sum_k \langle e_k | Q | e_k \rangle = \text{Tr} Q = m.
\]

Let \(d\lambda = \sum_k \delta_{\lambda_k}\) and apply Lemma 9.10 to \(d\mu = \Phi^2 d\lambda = \sum_k \Phi_k^2 \delta_{\lambda_k}\), to obtain \(r\) intervals \(I_k\) with \(\eta = C_x \epsilon^2, \kappa = C_y \epsilon^3\), and \(r \leq 2\kappa^{-1}\). We define \(x, y, C_x, C_y\) and \(z, C_z\) later. Note that \(f^2 d\lambda := \sum_k f_k^2 \delta_{\lambda_k}\) is dominated by \(d\mu\) which is how we will get estimates that depend only on \(m = \|\Phi\|_{L^2(d\lambda)}^2 = \mu([0, 1])\). Let \(S\) denote the set of \(\lambda_k\) that are in \(\bigcup_j I_j\). Then

\[
\|f \chi_{[0,1] \setminus S}\|_{L^2(d\lambda)} \leq \left(\frac{4\eta}{\kappa} \mu([0, 1])\right)^{1/2} \leq \left(\frac{4\eta m}{\kappa}\right)^{1/2}.
\]

Once these are done, let \(A_j = \chi_{I_j} (J) P\mathcal{V}_1\) with polar decomposition \(U_j Q_j\), then for \(a = C_x \epsilon^2\), set \(K_j = E_{[a, 1]}(Q_j)\). Then let \(L\) be the orthogonal sum \(\bigoplus_j (A_j K_j(\mathcal{V}_1))\). The space \(L\) will approximately satisfy the conditions that we want the range of \(P\) to satisfy and so by Lemma 4.6 we will have the result.
The bound \( \| (1 - P_L)P_{V_1} \| \leq (4\eta m/\kappa)^{1/2} + (2/\kappa)^{1/2} a \) follows because we only removed some subintervals from [0, 1] that have small contribution to the norm of a vector in \( P_{V_1} \). A little more specifically, \( \| A_j K_j - A_j \| \leq a \) and \( \sum_j A_j = (\sum_j \chi_{I_j}(J)P_{V_1} \) is approximately \( \chi_{[0,1]}(J)P_{V_1} = P_{V_1} \), because of the inequality in \([12]\).

The second inequality \( \| [J, P_L] \| \leq \kappa/2 \) is equivalent to \( \| (1 - P_L)JP_L \| \leq \kappa/2 \). For \( f \in \mathcal{L} \), write \( f = \sum_j t_j A_j K_j f^j \), where \( f^j \in \mathcal{V}_1 \) with \( \sum_j t_j^2 = 1 \), \( |A_j K_j f^j| = 1 \). Because \( K_j \leq P_{V_1} \) we can pick \( f^j \in \mathcal{V}_1 \) such that \( K_j f^j = f^j \). Therefore, \( f = \sum_j t_j A_j f^j \)

\[
|f^j| \leq a^{-1}|f|
\]
because \( 1 = |A_j K_j f^j| = |Q_j E(a,1)(Q_j)f^j| \).

We have to show that \( \| (1 - P_L)Jf \| \) is small. The ranges of the \( A_j \) are orthogonal, because \( A_j \leq \chi_{I_j}(J) \). The lengths of the \( I_j \) are at most \( \kappa \), so the terms \( A_j K_j f^j \in \mathcal{L} \) are almost eigenvectors of \( J \). So, \( \mathcal{L} \) is the orthogonal direct sum of almost eigenspaces.

In more detail, if \( c_j \) is the midpoint of the interval \( I_j \) then

\[
|J A_j f^j - c_j A_j f^j| = |(J - c_j)\chi_{I_j}(J)A_j f^j| \leq \frac{\kappa}{2}|A_j f^j|
\]
and so

\[
\begin{align*}
|\sum_j (1 - P_L)J t_j A_j f^j| &= |(1 - P_L)\left( \sum_j t_j J A_j f^j - \sum_j t_j c_j A_j f^j \right) | \\
&\leq \sum_j |t_j J A_j f^j - \sum_j t_j c_j A_j f^j| \leq \frac{\kappa}{2} \sqrt{\sum_j |t_j|^2 |A_j f^j|^2} = \frac{\kappa}{2} |f|.
\end{align*}
\]

The previous two inequalities will parallel very much to what Hastings’ method does, but this last inequality will not. A key difference is that while Hastings takes overlapping intervals with smooth cut-off functions, Szarek uses the disjointness of the intervals he chooses. More precisely, for \( \gamma = \frac{M}{\eta (L_0 - 1)} \), we get polynomials \( 0 \leq p_j \leq 1 \) of degree \( L_0 - 1 \) with \( 1 - \gamma \leq p_j \leq 1 \) on \( I_j \), \( 0 \leq p_j \leq \gamma \) on \( I_i \), \( i \neq j \) and \( \sum_j p_j = 1 \) on [0, 1].

We then estimate \( |f - \sum_j t_j p_j(J) f^j| \) for \( f \in \mathcal{L} \), because the \( p_j(J)f^j \) are in \( \bigoplus_{i=1}^{L_0-1} \mathcal{V}_i \) due to the degree restriction of the polynomials. This estimate is \( a^{-1} \left( 8\gamma^2/\kappa + 4\eta m/\kappa \right)^{1/2} \), which is seen as follows:

We know that

\[
f = \sum_j t_j A_j f^j = \sum_j t_j \chi_{I_j}(J) f^j.
\]

So,

\[
|\sum_j t_j \chi_{I_j}(J) f^j - t_j p_j(J) f^j|^2 = |\sum_j t_j \chi_{S(J)}(\chi_{I_j}(J) - p_j(J)) f^j|^2 \\
+ |\sum_j t_j \chi_{[0,1]\setminus S(J)}(\chi_{I_j}(J) - p_j(J)) f^j|^2 =: I_1 + I_2.
\]

For \( \lambda_k \in S \cap I_k \),

\[
\sum_j |\chi_{I_j}(\lambda_k) - p_j(\lambda_k)| \leq |\chi_{I_k}(\lambda_k) - p_k(\lambda_k)| + \sum_{j \neq k} |\chi_{I_j}(\lambda_k) - p_j(\lambda_k)|.
\]
Because $1 - \gamma \leq p_k \leq \chi_{I_k} = 1$ on $I_k$ and $\sum_{j \neq k} p_j = 1 - p_k$ maps into $[0, \gamma]$ on $I_k$, so for $\lambda \in I_k$,

$$\sum_j |\chi_{I_j}(\lambda) - p_j(\lambda)| \leq 2\gamma.$$ 

Writing $f^j = \sum_k f^j_k e_k$ gives

$$I_1 = \left| \sum_{\lambda_k \in S} \sum_j (\chi_{I_j}(\lambda_k) - p_j(\lambda_k)) t_j f^j_k e_k \right|^2 = \sum_{\lambda_k \in S} \left| \sum_j (\chi_{I_j}(\lambda_k) - p_j(\lambda_k)) t_j f^j_k \right|^2 \leq \sum_{\lambda_k \in S} \left( \max_j |t_j f^j_k| \sum_j |\chi_{I_j}(\lambda_k) - p_j(\lambda_k)| \right)^2 \leq (2\gamma)^2 \sum_{j,k} |t_j f^j_k|^2 = (2\gamma)^2 \sum_j |t_j f^j|^2 \leq (2\gamma)^2 r a^{-2} |f|^2.$$

Also,

$$I_2 = \left| \sum_j \chi_{[0,1]\setminus S}(J)(\chi_{I_j}(J) - p_j(J)) f^j \right|^2 \leq \sum_{\lambda_k \not\in S} \left| \sum_j (\chi_{I_j}(\lambda_k) - p_j(\lambda_k)) t_j f^j_k \right|^2 \leq \sum_{\lambda_k \not\in S} \sum_j |p_j(\lambda_k) t_j f^j_k|^2 \leq \sum_{\lambda_k \not\in S} \sum_j |t_j f^j_k|^2 = \sum_j |t_j|^2 \|f^j \chi_{[0,1]\setminus S}\|_{L^2(\lambda)}^2 \leq \frac{4\eta m}{\kappa} a^{-2} |f|^2.$$

So,

$$|f - \sum_j t_j p_j(J) f^j| \leq a^{-1} \left( (2\gamma)^2 r + \frac{4\eta m}{\kappa} \right)^{1/2} |f|.$$

Now, we can pick $\epsilon_1$ to be any value greater than $\epsilon_4 + 10 \max(\epsilon_3, \epsilon_5)$, where $\epsilon_3 = \left( \frac{4\eta m}{\kappa} \right)^{1/2} + (2/\kappa)^{1/2} a$, $\epsilon_4 = \frac{\epsilon}{2}$, and $\epsilon_5 = a^{-1} \left( 8\gamma^2 / \kappa + 4\eta m / \kappa \right)^{1/2}$. We have $\eta = C_x \epsilon^x, \kappa = C_y \epsilon^y, a = C_z \epsilon^z$. Also,

$$C_t \epsilon^l - 1 \leq L_0 - 2 \leq C_t \epsilon^l.$$

Because $l$ will be negative, we have

$$\epsilon < \left( \frac{C_t}{L_0 - 2} \right)^{1/|l|}.$$

We choose $x = 6, y = 1, z = 3/2, l = -9$ optimally. We then choose (likely not optimally) $C_x = 1/m, C_y = 2/11, C_z = 1, C_l = M m \sqrt{2}$. Then for $\epsilon \leq 1$, pick

$$\epsilon_1 := \left( \frac{1/11 + 10 \sqrt{11}(1 + \sqrt{2})}{11} \right)^{1/9} \epsilon.$$

The condition $\kappa > 8\eta$ translates into $\epsilon < \sqrt{m/44}$ which is satisfied if $L_0 > 44^{9/5} \sqrt{2} m^{-4/5} + 2$. Also, $\epsilon_1 < 83.4 \left( \frac{m M}{11^2 \sqrt{2}} \right)^{1/9}$ if $L_0 \geq \sqrt{2} m M + 2$.

---

Note: We could have obtained $(2\gamma)^2 a^{-2} |f|^2$ as an estimate of $I_1$ which gives a final exponent of $2/17$. This is only a slight improvement of the exponent gotten in [17] of $1/9$. 

29
10 Estimates, Putting it all Together

In this section we explore how the various constants discussed previously can be chosen to get the best possible decay in Lin’s theorem. There are cases where one might not want the best possible estimates, because perhaps picking a different rate allows one to not use Lemma 7.1 but instead use Szarek’s completely constructive result discussed in the previous section.

From $A, B$ self-adjoint with $\|A\|, \|B\| \leq 1, \|[A, B]\| \leq \delta$ we get (by applying Lemma 6.1) $H$ such that $\|H\| \leq 1, \|[H, B]\| \leq \text{Const.}\delta, \|[A - H]\| \leq \text{Const.}\delta/\Delta$. We get $[H', B'] = 0$ with $\|B - B'\| \leq 2/n_{cut} \sim 2\Delta^{\gamma_1}$ for $0 < \gamma_1 < 1$ and $\|H' - H\| \leq 2\epsilon_2 \leq 2L^{-\gamma_2}E(\frac{1}{\epsilon_2})$, where $L \sim \frac{2}{n_{cut}\Delta} \sim 2\Delta^{\gamma_1-1}$. So, writing $\Delta = \delta^{\gamma_0}, 0 < \gamma_0 < 1$, we get $2L^{-1} \sim \Delta^{1-\gamma_1} = \delta^{\gamma_0(1-\gamma_1)}$ so

$$\|A - H'\| \leq \text{Const.}\delta/\Delta + 2\epsilon_2 \leq \text{Const.}\delta^{1-\gamma_0} + \delta^{\gamma_0(1-\gamma_1)}\gamma_2 E_0(1/\delta) \quad (13)$$

and

$$\|B - B'\| \leq 2/n_{cut} \leq \text{Const.}\delta^{\gamma_0\gamma_1}. \quad (14)$$

We want to minimize these, noting that $\gamma_2$, which comes from Lemma 2, is the only constant that we cannot choose. Setting the three exponents equal we get $\gamma_0 = \frac{1}{1+\gamma_1}$ and $\gamma_1 = \frac{\gamma_2}{1+2\gamma_2}$, so the common value of the exponents will be chosen to be $\gamma = \frac{\gamma_2}{1+2\gamma_2}$.

Note that this method cannot give the optimal exponent of $\frac{1}{2}$ (only potentially $\frac{1}{2} + \epsilon$ if we are allowed to take $\gamma_2$ arbitrarily large). This is partly because of the averaging that we do to make $H$ “finite range” as expressed in the title of Section II.A of \[\Pi\], as a consequence of the $\delta/\Delta$ factor.

If we were able to remove that factor (by improving the result or starting in a special case of $A$ then we would then have to compare $\gamma_0(1 - \gamma_1)\gamma_2$ with $\gamma_0\gamma_1$ which gives $\gamma_1 = \frac{\gamma_2}{1+2\gamma_2}$. Because we need to have $A$ finite range of distance $\Delta$ with respect to $B$, we see that $\|[A, B]\| \leq \text{Const.}\Delta$. This gives $\delta \leq \text{Const.}\delta^{\gamma_0} \text{so } 0 < \gamma_0 \leq 1$. Because we do not need to worry about bounding $A - H$ (because we are starting with $A = H$) we get

$$\|A - H'\| \leq \delta^{\gamma_0(1-\gamma_1)}\gamma_2 E_0(1/\delta) \quad (15)$$

and

$$\|B - B'\| \leq 2\delta^{\gamma_0\gamma_1}. \quad (16)$$

So, we see that the rate we get is $\gamma = \frac{\gamma_2}{1+2\gamma_2}$, which is an improvement. The value of $\gamma_2$ is 1 to then get the optimal exponent $\gamma = 1/2$. If we have $\gamma_2 = 1/9$ as in Szarek’s result, then $\gamma = 1/10$.

Note that Davidson in the proof of the equivalence of $(Q)$ and $(Q’)$ showed that Lin’s theorem implies that we can pick $\gamma_2 = 1/2$, which is an estimate that gives Lin’s theorem with $\gamma = 1/4$. If one shows that we can pick $\gamma_2$ arbitrarily large, this shows that there is no loss in asymptotic dependence in the equivalence between $(Q)$ and $(Q’)$.
11 Smooth Partitions and Hastings’ Approach

The method of proof of Lemma 2 in \([1]\) in getting Item 3 is to utilize smooth step functions that overlap along the lines of Section 8. So, we define the general profile function that we use. Let \(F\) be some smooth strictly decreasing function on \([0, 1]\) with all derivatives at 0 and 1 equal to zero, \(F(0) = 1, F(1) = 0,\) and \(F(x) + F(1 - x) = 1\) for all \(x \in [0, 1].\) Define \(F_{\omega_0}^{r,w}(\omega)\) for \(\omega_0 \in \mathbb{R}, w > 0, r \geq 0\) by translation

\[
F_{\omega_0}^{r,w}(\omega) = F_0^{r,w}(\omega - \omega_0)
\]

where \(F_0^{r,w}\) is even, identically 1 on \([w, w + r]\), and equal to \(F((x - r)/w)\) on \([0, w]\). So \(F_{\omega_0}^{r,w}\) is identically equal to 1 on an interval centered at \(\omega_0\) of radius \(r\) and smoothly decreases smoothly to zero within the width \(w\). (See Figure 1 below.) The symmetry property of \(F\) implies that if \(I_i = (\omega(i) - \kappa, \omega(i) + \kappa)\) are nonconsecutively disjoint intervals whose union contains \([-1, 1]\) then \(\{F_{\omega(i)}^{0,\kappa}(\omega)\}\) forms a partition of unity covering \([-1, 1]\). See Figure 2 below.

\[\text{Figure 2: The graph of } F_{\omega_0}^{r,w}(\omega).\]

Note that although we choose \(F_{\omega_0}^{r,w}\) to be smooth, we could pick it to be smooth enough, by picking it to be piecewise polynomial (constructed by convolution of step functions then normalized so it is a partition of unity). In this case, terms like “faster than any polynomial” will become polynomial growth of a certain order. The modification changes \(\delta^\gamma E(1/\delta)\) to \(\delta^\gamma + \epsilon\) and how smooth that \(F\) needs to be can be seen at the very end of the proof of the proposition where we put together the estimates for \(T(l)\) and \(G(l)\). This then eliminates the very slowly growing functions \(E(1/\delta), F, T, G\) with much more explicit power functions.

Returning to the construction, fix the intervals \(I_i\) now. Below, we choose \(F(L)\) to grow slower than any power of \(L\) and so that our Lieb-Robinson estimates work out. For \(\beta_1 \in (0, 1),\) let \(n_{\text{win}} = \left\lceil L^{\beta_1} / F(L) \right\rceil\) writing \(\omega(i) = -1 + ki,\) where \(\kappa = 2/n_{\text{win}},\) for \(i = 0, \ldots, n_{\text{win}}.\) We then let \(I_i\) to be the interval of radius \(\kappa\) centered at \(\omega(i).\) So, the graphs of the \(F_{\omega(i)}^{0,\kappa}\), are centered at \(\omega(i)\) with radius \(\kappa\) and have no “flat part” and form a partition of unity covering the interval \([-1, 1].\)

**Definition 11.1.** Let \(S_1\) be a matrix whose columns \(v_1, \ldots, v_d\) are an orthonormal basis of \(V_1.\) Let \(\tau_i = [F_{\omega(i)}^{0,\kappa}(J)]S_1\) and \(Z_i = E_{[0, \lambda_{\text{min}}]}(\tau_i^* \tau_i).\) We define \(\lambda_{\text{min}}\) later. Then let \(X_i\) be the range of \(\tau_i(1 - Z_i).\)
For $j = 0, 1$, the functions $\hat{F}_0^{j,w}(t)$ are step functions that are equal to 1 on $[-jw, jw]$ and zero outside $(-(j + 1)w, (j + 1)w)$, where we imagine $w$ being large.

Now, note that
\[ G_\beta \text{ polynomial of } L_1 \leq 1 \text{ and } jw, jw \]

Now, set
\[ f \]
\[ \text{and we also assume } G \geq 2. \]

Now, set
\[ S(L) = \int_{|k| \geq L^{-1}/(e^2 n_{\text{win}})} |\hat{F}_0^{0,1}(k)| dk + \| \hat{F}_0^{0,1} \|_{L^1(\mathbb{R})} e^{-\frac{L^{-1}}{2}} \]
\[ T(l) = 2 \int_{|k| \geq L^{-1}/(e^2 n_{\text{win}})} |\hat{F}_0^{1,1}(k)| dk + 3 \| \hat{F}_0^{1,1} \|_{L^1(\mathbb{R})} e^{-\frac{l}{2}}. \]

Note the implicit dependence of $n_{\text{win}}$ on $L$ in the definition of $S(L)$. The intent of these definitions and the following estimates is to take advantage of the Lieb-Robinson estimates gotten in Section 6.

If $f \in H^n(\mathbb{R})$ and then because
\[ \int_{|k| \geq c} \frac{1}{1 + |k|^{2n}} dk = \int_{|k| \geq 1} \frac{c}{1 + c^2 |k|^{2n}} dk \leq \frac{2}{c^{2n-1}} \int_1^\infty \frac{1}{k^{2n}} dk = \frac{2}{(2n - 1)c^{2n-1}} \]
we have
\[ \int_{|k| \geq c} |\hat{f}(k)| dk \leq \sqrt{\int_{|k| \geq c} \frac{1}{1 + |k|^{2n}} dk} \sqrt{\int_{\mathbb{R}} |(1 + |k|^{2n})| \hat{f}(k)|^2 dk} \]
\[ \leq \frac{1}{c^{n-1/2}} \sqrt{\frac{2}{(2n - 1)}} \left( \|f\|_{L^2} + \|\hat{f}(n)\|_{L^2} \right) = \text{Const.} \frac{1}{c^{n-1/2}} \left( \|f\|_{L^2} + \|f(n)\|_{L^2} \right). \]
By picking \( F \) such that \( |(F_{0}^{1,1})'| \leq 2 \) and because we know that \( \text{supp} F_{0}^{1,1} \subset [-2,2] \) and \( \|F_{0}^{1,1}\|_{L^\infty} = 1 \) we can estimate
\[
\|\hat{F}_{0}^{1,1}\|_{L^1} \leq \sqrt{\int_{\mathbb{R}} \frac{1}{1 + |k|^2} dk} \sqrt{\int_{\mathbb{R}} (1 + |k|^2)|\hat{F}_{0}^{1,1}(k)|^2 dk} \\
\leq \text{Const.} \left( \|\hat{F}_{0}^{1,1}\|_{L^2} + \left\| (\hat{F}_{0}^{1,1})' \right\|_{L^2} \right) \leq \text{Const.} \left( \sqrt{4 + 2\| (\hat{F}_{0}^{1,1})' \|_{L^1}} \right) = \text{Const.}
\]

### 12 The use of Lin’s theorem in Lemma 6.1

The following result from \cite{[1]} is where Lin’s theorem is used.

**Lemma 12.1.** For any \( \epsilon \in (0, 1) \), there is an \( \delta > 0 \) such that if \( A, B \in M_n(\mathbb{C}) \) are self-adjoint with \( \|A\|, \|B\| \leq 1 \) and \( \|[A, B]\| \leq \delta \) then there is a projection \( P \) such that \( \|P, B\| \leq \epsilon \) and \( E_{[-1,1/2]}(A) \leq P \leq 1 - E_{[1/2,1]}(A) \).

**Remark 12.2.** Although we will only use this result for only one \( \epsilon \in (0, 1) \), the proof uses the “For all \( \epsilon > 0 \) there exists a \( \delta > 0 \)” aspect of Lin’s theorem for only one \( \epsilon \in (0, 1/22) \).

This proof is a dramatic simplification of the proof in \cite{[1]} using a simplified form of an argument used in Theorem 3.2 of \cite{[5]}.

**Proof.** Let \( A', B' \in M_n(\mathbb{C}) \) be any commuting self-adjoint matrices. Let \( E = E_{[-1,1/2]}(A), G = E_{[-1,1/2]}(A') \) and set \( P' = E_{(-\infty,0)}(A') \). We have
\[
\|[P', B]\| = \|[P', B - B']\| \leq 2\|B - B'\|.
\]

Also, by the Davis-Kahan theorem,
\[
\|EP'^{\perp}\| = \|E_{[-1,-1/2]}(A)E_{[0,\infty)}(A')\| \leq 2\|A - A'\|
\]
and
\[
\|P'G^{\perp}\| = \|E_{(-\infty,0)}(A')E_{[1/2,1]}(A)\| \leq 2\|A - A'\|.
\]

Then we apply Lemma 4.6\footnote{This is a reference to a specific lemma, possibly in a previous section.} to get that there is a projection \( P \) such that \( E \leq P \leq G \) with \( \|P - P'\| \leq 10\|A - A'\| \) and
\[
\|[P, B]\| \leq \|[P - P', B]\| + \|[P', B]\| \leq 20\|A - A'\| + 2\|B - B'\|.
\]

Now by Lin’s theorem, there is a \( \delta > 0 \) such that if \( \|[A, B]\| < \delta \) we can pick \( A', B' \) such that \( \|A - A'\|, \|B - B'\| \leq \epsilon/22 \) so that the projection \( P \) satisfies the conditions of the lemma. \( \square \)

### 13 Many subspaces

Recall the spaces \( X_i \) from above. Spaces that are indexed with nonconsecutive indices are orthogonal, but otherwise we have no control over the orthogonality of these spaces. As mentioned at the end of Section \ref{section:many_subspaces} what we want to do is be able to write \( x \in \)
$X_0 + \cdots + X_{n_{\text{win}}-1} = \mathcal{X}$ as $\sum_{i=0}^{n_{\text{win}}} x_i, x_i \in \mathcal{X}_i$ with $|v|^2$ comparable to $\sum_i |x_i|^2$. We cannot guarantee that this is possible, so we attempt to work around it.

Make spaces $\mathcal{R}_i$ of the same dimension of $\mathcal{X}_i$ and define $A : \bigcup \mathcal{R}_i \to \bigcup \mathcal{X}_i$ to be the natural identification. We then define $\mathcal{R}$ as the (exterior) direct sum $\bigoplus \mathcal{R}_i$ and extend $A : \mathcal{R} \to \mathcal{X}$. Note the key property that $A|\mathcal{R}_i$ is an isometry. Let $x \in \mathcal{X}$ with $r_i \in \mathcal{R}_i$ such that $x_i = Ar_i$. What we are looking for is a way to make $|x|^2 = |\sum_i Ar_i|^2 = |A(\sum_i r_i)|^2$ comparable to $\sum_i |x_i|^2 = \sum_i |r_i|^2 = |\sum_i r_i|^2$.

Just as before with $\mathcal{F}_{\omega}(J)\mathcal{V}_1$, in order to obtain a lower bound for $|A(\sum_i r_i)|^2$ we involve ourselves with finding a way to approximately “cut” the subspace $E_{\lambda_{\text{small}}}(A^*A)$ out of our subspace $\mathcal{R}$. We will form a subspace $\mathcal{U}$ approximately out of spaces $\mathcal{N}_i$ so that $E_{\lambda_{\text{small}}}(A^*A)$ is approximately encompassed by the span of the $\mathcal{N}_i$ and then choose $\mathcal{W} = \mathcal{A}U^{-1}$, which will be a subspace of $\mathcal{X}$. Because we would like to use the fact that the $\mathcal{X}_i$ are approximate eigenspaces, we might hope that $E_{\lambda_{\text{small}}}(A^*A)$ breaks down into an orthogonal sum of subspaces $\mathcal{N}_i$ of $\mathcal{R}_i$. It is not clear that this is possible, but what we can show is that by merging long chains of the $\mathcal{R}_i$ together into subspaces $\mathcal{Y}_i$ that have considerable consecutive overlap we can get $E_{\lambda_{\text{small}}}(A^*A)$ to be approximately the span of $\mathcal{N}_i \subset \mathcal{Y}_i$. Because we do not expect there to be orthogonality (or even linear independence) between the $\mathcal{N}_i$, we need to concern ourselves with representing each vector projected onto by $E_{\lambda_{\text{small}}}(A^*A)$ in a manageable way. We now proceed to the details.

Let $\rho = A^*A$. Then $\rho$ is positive and block tridiagonal: if $r_k \in \mathcal{R}_k$ for all $k$, $Ar_k \in \mathcal{X}_k$ so $(r_i, pr_j) = (Ar_i, Ar_j) = 0$, if $|i - j| > 1$. Choose a basis for $\mathcal{X}$ by picking an orthonormal basis $\{v^I_j\}_j$ of $\mathcal{X}_i$ then removing elements of $\bigcup_j \{v^I_j\}_j$ so that it is a basis of unit vectors for $\mathcal{X}$. Then $A$ is tridiagonal with respect to this basis. Moreover, because $A|\mathcal{R}_i$ is an isometry, the blocks of $\rho$ on the diagonal are identity matrices.

Now, we merge some of the spaces $\mathcal{R}_i$ to attempt to take advantage of the block tridiagonal nature of $\rho$. As a notational convenience, if $I$ is an interval, then let $\mathcal{R}_I = \bigoplus_{i \in I} \mathcal{R}_i$. Define $\mathcal{X}_I$ likewise. Let $n_b \sim L^{\beta_0}$ be the number of “superblocks” that we form most of which have length $l_b$. That is, we define $k_b = \lfloor (n_{\text{win}} + 1)/l_b \rfloor - 1$ and then:

$$n_b = \left\{ \begin{array}{ll} k_b & \text{if } k_b \text{ is odd,} \\ k_b - 1 & \text{if } k_b \text{ is even.} \end{array} \right.$$  

This is so that $n_b$ is always odd.

For $1 \leq i \leq n_b - 1$, let $\mathcal{Y}_i$ be the $i$th “superblock” defined by $\mathcal{Y}_i = \mathcal{R}_{[i(i-1)/l_b, (i+1)/l_b)}$. See Figure 3 below. Let $\mathcal{R}_j$ be a projection in $\mathcal{R}$ onto $\mathcal{R}_j$, and for $1 < i \leq n_b - 1$ let $Y_i$ project onto $\mathcal{Y}_i$, $Y'_i$ project onto $\mathcal{R}_{[i(i-3)/l_b, (i+3/4)/l_b)}$, and $Y''_i$ project onto $\mathcal{R}_{[(i-1)/l_b, (i+1/2)/l_b)}$. Note that $Y'_{i-1}Y'_i = Y'_iY'_{i+1}$ projects onto (a “left” subspace of the image of $Y'_i$) $\mathcal{R}_{[(i-3)/l_b, (i-1)/l_b)}$ and $Y'_{i+1}Y'_i = Y'_iY'_{i+1}$ projects onto (a “right” subspace of the image of $Y'_i$) $\mathcal{R}_{[(i+1)/l_b, (i+3/4)/l_b]}$. Note that there are $l_b/2$ many blocks between the images of $Y''_{i-1}Y''_i$ and $Y''_{i+1}Y''_i$. See Figure 4 below.

Note that there are two distinguishable subspaces $\mathcal{Y}_i$ that we address now, namely $\mathcal{Y}_1$ and $\mathcal{Y}_{n_b}$. We define these spaces and their respective projections $Y_i, Y'_i, Y''_i$ to address the issue that $\mathcal{Y}_1$ and $\mathcal{Y}_{n_b}$ only intersect the other intervals on one side. The first subspace $\mathcal{Y}_1$ is a full length interval that only intersects other $\mathcal{Y}_i$ on its right side. And the last subspace $\mathcal{Y}_{n_b}$ has not yet been specified, but we do that now. The idea is that in order to apply the Lieb-Robinson estimates, we want all the intervals to have length at least $2l_b$ (and at most $3l_b$), so we make $\mathcal{Y}_{n_b} = \mathcal{R}_{[(n_b-1)/l_b, n_{\text{win}}]}$. For these two
Figure 3: The subspaces $X_i, R_i, Y_i$ are illustrated above with 33 blocks, $l_b = 3$, $n_{win} = 32$, $n_b = 8$. In this case we could define $Y_{10}$ to be the last six blocks, but not if $27 \leq n_{win} \leq 31$ in which case we would have no guarantee of a minimal length of $Y_{10}$. Note the distinction between the first and last of the $Y_i$. Note that the subspaces $X_i$ are nonconsecutively orthogonal (and not even necessarily linearly independent), while the $R_i$ are orthogonal. Also, the subspaces $Y_i$ are nonconsecutively orthogonal and are displayed to show this.

Figure 4: Typical consecutive projections $Y_i, Y_i', Y_i''$ are illustrated above with $l_b = 8$. 

Note the distinction between the first and last of the $Y_i$. Note that the subspaces $X_i$ are nonconsecutively orthogonal (and not even necessarily linearly independent), while the $R_i$ are orthogonal. Also, the subspaces $Y_i$ are nonconsecutively orthogonal and are displayed to show this.
subspaces we now define the respective projections \( Y_i, Y_i', Y_i'' \) as we did above.

For all \( i \), \( Y_i \) projects onto \( Y_i' \), \( Y_i' \) projects onto \( \mathcal{R}_{[0,(1+3/4)l_b]} \), \( Y_i'' \) projects onto \( \mathcal{R}_{[0,(1+1/2)l_b]} \), \( Y_{n_b}' \) projects onto \( \mathcal{R}_{[(n_b-3/4)l_b,n_{win}]} \), \( Y_{n_b}'' \) projects onto \( \mathcal{R}_{[(n_b-1/2)l_b,n_{win}]} \).

The point is that the \( Y_i'' \) form a resolution of the identity and there are \( 4l_b \) many blocks between the blocks projected onto by \( Y_i'' \) and the blocks not projected onto \( Y_i' \). To avoid multiple cases when dealing with these projections, let \( Y_0 \) project onto \( \mathcal{R}_{[0,l_b]} \), \( Y_0' \) project onto \( \mathcal{R}_{[0,3l_b/4]} \), \( Y_{n_b+1} \) project onto \( \mathcal{R}_{[n_b,4n_{win}]} \), and \( Y_0' \) project onto \( \mathcal{R}_{[(n_b+1/4)l_b,n_{win}]} \), so the “left” and “right” subspaces are well-defined for all \( i = 1, \ldots, n_b \).

### 14 Construction and Properties of the spaces \( \mathcal{N}_i \)

We construct spaces \( \mathcal{N}_i \) that essentially encompass the small eigenvalue eigenvectors of \( \rho \), are subspaces of \( Y_i' \), and have various properties that we explore in this section. The major modifications in this section of the proofs from [1] were suggested in [3].

**Definition 14.1.** Let \( \rho_i = Y_i' \rho Y_i' \) be \( \rho \) projected onto the range of \( Y_i' \) and \( \hat{B}_i \) a position operator for subspaces \( \{ R_j \} \) of \( Y_i \) such that \( \hat{B}_i = -I \) restricted to the range of \( Y_i' \), is \( I \) restricted to the range of \( Y_i' \), and linearly interpolates as multiples of the identities on blocks \( R_j \) with \( \hat{B}_i \) on \( R_j \) being \( \frac{2}{l_b/2+1} [j - (i + 1/4)l_b] + 1 \) \( I \) for \( i - 1/4l_b \leq j < (i + 1/4)l_b \).

Let \( \chi \in (0,1) \) be some constant that we will pick later.

**Lemma 14.2.** For \( l_b \) large enough, there exist subspaces \( \mathcal{N}_i \) (projected onto by \( \mathcal{N}_i \)) such that \( \mathcal{N}_i \leq Y_i' \), where

\[
E_{[0,G(l_b)/l_b]}(\rho_i) \leq \mathcal{N}_i \leq Y_i' - E_{[2G(l_b)/l_b,\infty]}(\rho_i)
\]

and

\[
\| \| \mathcal{N}_i, \hat{B}_i \| \| \leq 1 - \chi < 1.
\]

If \( \delta \) is the constant from Lemma 12.1 for \( \epsilon = 1 - \chi \), the condition on \( l_b \) of \( G(l_b) > 16C_{\mathcal{F}_0,1}/\delta \) can be taken as “large enough”.

**Proof.** The idea is to apply Lemma 12.1 to \( \hat{B}_i \) and some function of \( \rho_i \).

Because \( \chi < 1 \) is independent of \( l_b \) and that we require \( G(l_b) \) go to infinity as \( l_b \) does, we know that eventually \( G(l) > 16C_{\mathcal{F}_0,1}/\delta \), where \( \delta \) is a chosen value for \( \epsilon = 1 - \chi \) from Lemma 12.1.

Because \( \rho_i \) and \( \hat{B}_i \) both act on \( Y_i' \), \( \rho_i \) is block tridiagonal, and \( \hat{B}_i \) is a direct sum of multiples of identity matrices, with differences between consecutive multiples at most \( \frac{4}{l_b+2} \), we see that \( \| \| \rho_i, \hat{B}_i \| \| \leq \frac{8}{l_b+2} \). Let \( E_{1/2,1}(f_{l_b}(x)) = 1 - 2\mathcal{F}_0^{G(l_b)/l_b}G(l_b)/l_b(x) \). Then

\[
E_{[-1,-1/2]}(f_{l_b}(\rho_i)) \geq E_{[0,G(l_b)/l_b]}(\rho_i)
\]

and

\[
E_{[1/2,1]}(f_{l_b}(\rho_i)) \geq E_{[2G(l_b)/l_b,\infty]}(\rho_i).
\]
By equation (18) and by Proposition 5.6 we get

\[ \|f_i(\rho_i), \hat{B}_i\| \leq 2C_x f_{2G(l_b)/l_b} \|\rho_i, \hat{B}_i\| \leq \frac{16l_b}{(l_b + 2)G(l_b)} C_x f_{1,1} < \frac{16C_x f_{1,1}}{G(l_b)}. \]

So, for \( l_b \) large enough we can apply Lemma [12.1] to obtain \( N_i \) such that

\[ E_{[0, G(l_b)/l_b]}(\rho_i) \leq E_{[-1, -1/2]}(f_i(\rho_i)) \leq N_i \perp E_{[1/2, 1]}(f_i(\rho_i)) \geq E_{[2G(l_b)/l_b, \infty]}(\rho_i) \]

and \( \|\hat{N}_i, \hat{B}_i\| < 1 - \chi. \)

Remark 14.3. Instead of requiring \( l_b \) to be large enough, we could make \( G(l_b) \) have a lower bound that is large enough. However, how large would be left undetermined due to the statement of Lin’s theorem applied to get Lemma [12.1]. Similar remarks could be made elsewhere.

Definition 14.4. Let \( N^e = \sum_{i \text{ even}} N_i, N^o = \sum_{i \text{ odd}} N_i. \) Note that these are projections because the first item of the next result shows that the \( N_i \) are nonconsecutively orthogonal.

We are interested in controlling the orthogonality of the spaces \( N_i \), the elements of the \( N_i \) as elements of eigenspaces of \( \rho \), and the representations of eigenvectors (for small eigenvalues) of \( \rho \) by the spaces \( N_i \). Most of the following results are inherent in [3], but the last inequality in Item 3 is by Hastings in [3].

Lemma 14.5. For \( l_b \) large enough defined by and for the \( N_i \) satisfying Lemma [14.2] we have the following properties:

1. The spaces \( N_i \) are nonconsecutively orthogonal.
2. For any \( v \in N_i, 0 \leq (v, \rho v) \leq \frac{2G(l_b)}{l_b} |v|^2. \)
3. For \( v \) in the range of \( E_{[0, 1/l_b]}(\rho) \), there are \( n_i \in N_i \) such that

\[ |v - \sum_i n_i| \leq T(l_b) \sqrt{\rho} |v|, \]

\[ \sum_i |n_i|^2 \leq |v|^2, \]

and

\[ \sum_{i \text{ even}} |n_i|^2 \leq |N^e v||v| + T(l_b) \sqrt{\rho} |v|^2. \]

4. \( \|Y_{i+1} N_i Y_{i-1}^\dagger\| \leq 1/2 - \chi/2. \)

Proof. Item 1 follows because \( N_i \subset \mathcal{Y}_i \) and the \( \mathcal{Y}_i \) are nonconsecutively orthogonal by construction.

Item 2 is trivial once we remember that \( \rho_i = Y_{i}^\dagger \rho Y_i^\dagger, N_i \leq E_{[0, 2G(l_b)/l_b]}(\rho_i), \) and \( N_i \leq Y_i^\dagger \) so \( (v, \rho v) = (v, \rho v). \)

We prove Item 4. Because \( Y_{i-1} Y_i^\dagger = E_{(-1)}(\hat{B}_i), Y_{i+1} Y_i^\dagger = E_{(1)}(\hat{B}_i), \) and \( N_i \leq Y_i^\dagger, \) by Proposition 5.4 we obtain

\[ \|Y_{i+1} N_i Y_{i-1}^\dagger\| = \|Y_{i+1} N_i Y_{i-1}^\dagger N_i Y_{i+1} Y_i^\dagger\| = \|E_{(1)}(\hat{B}_i) N_i E_{(-1)}(\hat{B}_i)\| \leq \frac{\|N_i, \hat{B}_i\|}{2} \leq \frac{1 - \chi}{2}. \]
We now prove Item 3. Recall that \( G \geq 2 \). Let \( D = \mathcal{F}_0^{G(l_b)/2l_b, G(l_b)/2l_b}(\rho) \), \( D'_i = \mathcal{F}_0^{G(l_b)/2l_b, G(l_b)/2l_b}(\rho_i) \). Note that \( 0 \leq D \leq E_{(0, G(l_b)/l_b)}(\rho) \) and \( 0 \leq D'_i \leq E_{(0, G(l_b)/l_b)}(\rho_i) \leq N_i \). We will apply a Lieb-Robinson estimate to \( D, D' \) to obtain the \( n_i \) with the desired properties.

Because \( G \geq 2 \) so that \( \mathcal{F}_0^{G(l_b)/2l_b, G(l_b)/2l_b}(t) = 1 \) for \( |t| \leq G(l_b)/2l_b \geq 1/l_b \), we have that for \( v \) in the range of \( E_{[0, 1/l_b]}(\rho) \), \( Dv = v \). If \( v_i = Y_i''v \), then \( v = \sum_i v_i \) is an orthonormal decomposition.

Let \( S' \) be the set of eigenvalues of \( \hat{B}_i \) on \( Y_i'' \) and \( S'' \) the set of eigenvalues of \( \hat{B}_i \) on \( Y_i'' \). Now, \( \rho_i = Y_i'' Y_i' \) is tridiagonal with respect to the eigenspaces of \( \hat{B}_i \), so \( \rho_i \) is finite range with distance less than \( \Delta = 5/(l_b + 2) \). Because \( \text{dist}(S'', \mathbb{R} \setminus S') = \frac{l_b}{4} \left( \frac{1}{l_b + 2} \right) \), we see that \( \text{dist}(S'', \mathbb{R} \setminus S')/\Delta \geq l_b/5 \). Hence using Corollary 6.10 along with (17), (19) and (22), we have

\[
\|(D - D'_i)Y_i''\| \leq T(l_b).
\]

So, set \( n_i = D'_i v_i \). Then because \( Dv = v \),

\[
|v - \sum_i n_i| = \left| \sum_i (Dv_i - D'_i v_i) \right| \leq T(l_b) \sum_i |v_i| \leq T(l_b) \sqrt{\sum_i |v|^2}
\]

and

\[
\sum_i |n_i|^2 = \sum_i |D'_i v_i|^2 \leq \sum_i |v_i|^2 = |v|^2.
\]

When restricting to just the even indices we get a different bound using the orthogonality of \( n_i = D'_i v_i \in \mathcal{N}_i \) for even \( i \). Recall that \( D \) is self-adjoint so:

\[
\sum_{i \text{ even}} |n_i|^2 \leq \sum_{i \text{ even}} |v_i|^2 = \left( v, \sum_{i \text{ even}} v_i \right) = \left( Dv, \sum_{i \text{ even}} v_i \right) = \left( v, \sum_{i \text{ even}} Dv_i \right)
\]

\[
\leq \left| \left( v, \sum_{i \text{ even}} D'_i v_i \right) \right| + \left| \left( v, \sum_{i \text{ even}} (Dv_i - D'_i v_i) \right) \right|
\]

\[
\leq \left| \left( v, \sum_{i \text{ even}} n_i \right) \right| + \sum_{i \text{ even}} |(v, Dv_i - D'_i v_i)| \leq \left| N^e v, \sum_{i \text{ even}} n_i \right| + T(l_b)|v| \sum_{i \text{ even}} |v_i|
\]

\[
\leq |N^e v| \sum_{i \text{ even}} n_i + T(l_b)\sqrt{\sum_i |v|^2} = |N^e v| \left( \sum_{i \text{ even}} |n_i|^2 \right)^{1/2} + T(l_b)\sqrt{\sum_i |v|^2}
\]

\[
\leq |N^e v| \sqrt{\sum_{i \text{ even}} |v_i|^2} + T(l_b)\sqrt{\sum_i |v|^2} \leq |N^e v||v| + T(l_b)\sqrt{\sum_i |v|^2}.
\]

\[\square\]

**Remark 14.6.** Although a similar result is true for \( \sum_{i \text{ odd}} |n_i|^2 \), we will not use it.

**Remark 14.7.** Note that because \( Y_i' \perp Y_{i+1}' \) the inequality \( \|Y_{i+1}N_i Y_{i-1}\| \leq 1/2 - \chi/2 \) is a slight improvement of the general property of that if \( P, Q \) are any projections then \( \|(1 - P)QP\| \leq 1/2 \).

This property follows by applying Jordan’s lemma to reduce to the 2-dimensional case where one has \( P \) projecting onto the first basis vector \( e_1 \), \( 1 - P \) projecting onto the second basis vector \( e_2 \), and \( Q \) projecting onto a vector \( v = \cos \theta e_1 + \sin \theta e_2 \). A calculation then shows that \( \|(1 - P)QP\| = |\cos \theta \sin \theta| = \frac{1}{2} |\sin 2\theta| \).

38
**Definition 14.8.** Fix some \( \eta \in (0, \chi/4) \). Let \( i \) be odd. Apply Jordan’s lemma to \( N_i \) and \( N^e \) to get an orthonormal basis \( \{n_s\} \) of \( N_i \) such that \( \langle N^e n_s, n_t \rangle = 0 \) if \( s \neq t \). Then let \( N'_i \) be the subspace of \( N_i \) generated by the \( n_s \) such that \( |N^e n_s|^2 \leq 1/2 + \eta \). Then let \( N'_i \) project onto \( N'_i \) and \( N' = \sum_{i \text{ odd}} N'_i \).

**Remark 14.9.** Note that \( N'_i \) is not defined for \( i \) even, so \( N' \) unambiguously projects onto the orthogonal sum of \( N'_i \) for \( i \) odd. We also have that \( N_j N^o N_i = N_j N^e N_i = N_j N' N_i = 0 \) for \( |i - j| \geq 3 \). By our definitions, if \( n'_i \in N'_i \) is a unit vector then we can express it as \( n'_i = \sum_s c_s n_s \), where the \( n_s \) are as above. Then

\[
|N^e n'_i|^2 = |\sum_s c_s N^e n_s|^2 = \sum_s |c_s N^e n_s|^2 \leq (1/2 + \eta) \sum_s |c_s|^2 = 1/2 + \eta
\]  

and consequently

\[
|(1 - N^e)n'_i|^2 \geq 1/2 - \eta.
\]

Note that if \( n_i \in N_i \oplus N'_i \) then \( n_i \) is expressed as linear combination of the \( n_s \) with \( |N^e n_s|^2 > 1/2 + \eta \) and hence the opposite inequalities hold:

\[
|N^e n_i| > \sqrt{1/2 + \eta}
\]

and

\[
|(1 - N^e)n_i| < \sqrt{1/2 - \eta}.
\]

See that the definition of \( N'_i \) is intended to remove a part of \( N_i \) that has large projection onto \( N_{i-1} \oplus N_{i+1} \). This is not quite enough control of the orthogonality to give a result like Equation (2), but, along with \( \|Y'_{i+1} N_i Y'_{i-1}\| < 1/2 - \chi/2 \), it will suffice for our purposes.

**Definition 14.10.** Let \( U \subset R \) be defined as the orthogonal complement of the span of \( \bigcup_{i \text{ even}} N_i \cup \bigcup_{i \text{ odd}} N'_i \). Then let \( W = AU \subset X \) and \( P \) project onto \( W \).

**Remark 14.11.** \( U^\perp \) is expressed as the direct sum \( \bigoplus_{i \text{ odd}} N_{i+1} \oplus N'_i \), because these subspaces are linearly independent.

We show that this is a direct consequence of the fact that \( \|N'_{i+2} N_{i+1} N'_i\| \leq 1/2 - \chi/2 \) and how when constructing \( N'_i \) we “cut out” the part of \( N_i \) that is almost parallel to the image of \( N^e \). Note that this linear independence is also a consequence of the first inequality in the following lemma. This does not show that all the subspaces \( N_i \) or \( N'_i \) are nonzero.

We show the linear independence as follows. Suppose that \( n_i \in N_i \) for \( i \) even and \( n_i \in N'_i \) such that

\[
\sum_{i \text{ even}} n_i = -\sum_{i \text{ odd}} n_i.
\]

By nonconsecutive orthogonality, Equation (27) shows that

\[
\sum_{i \text{ even}} |n_i|^2 = \sum_{i \text{ even}} |n_i|^2 = |\sum_{i \text{ odd}} n_i|^2 = \sum_{i \text{ odd}} |n_i|^2.
\]
However, applying $1 - N^c$ to Equation (27) also shows that for $i$ odd

$$0 = \left| \sum_{i \text{ even}} (1 - N^c) n_i \right|^2 = \left| \sum_{i \text{ odd}} (1 - N^c) n_i \right|^2.$$

Following the calculation below in Equation (29), we see that

$$|((1 - N^c) n_i, (1 - N^c) n_{i+2})| = |(Y'_{i+2} N_{i+1} Y'_{i} n_i, n_{i+2})| \leq \|Y'_{i+2} N_{i+1} Y'_{i}\| \|n_i\| n_{i+2}|.$$

So, by Equation (3)

$$0 = \left| \sum_{i \text{ odd}} (1 - N^c) n_i \right|^2 \geq (1 - 2(1/2 - \chi/2)) \left| \sum_{i \text{ odd}} (1 - N^c) n_i \right|^2 \geq \chi(1/2 - \eta) \left| \sum_{i \text{ odd}} n_i \right|^2.$$

This implies that $n_i = 0$ for $i$ odd and hence $n_i = 0$ for all $i$ by Equation (28).

**Lemma 14.12.** There are constants $C_1 = C_1(\chi, \eta) \geq 1$ and $\alpha = \alpha(\chi, \eta) \in (0, 1)$ such that for any $y_i \in \mathcal{Y}_i$, there are (unique) $n_j^i \in \mathcal{N}_j$ for $j$ even and $n_j'^i \in \mathcal{N}_j'$ for $j$ odd such that $U^\perp y_i = \sum_j n_j^i$ and $|n_j^i| \leq C_1 \alpha |i-j| |y_i|.$

Consequently, for $c_\alpha = (1 + \alpha + \alpha^{-1})$ and $C_2 = c_\alpha C_1, \|Y_j U Y_i\| \leq C_2 \alpha |i-j|.$

**Proof.** Note that for the proof of the first part, it will be important to consider different values of $|n_j^i|$ for the same unit vector $n_j^i/n_j'^i$. So, for the proof we will write $n_j^i = a_j^i m_j^i$, where $m_j^i$ is always a unit vector (even if $|a_j^i| = |n_j^i| = 0$).

We first prove the second result. Note that because $C_2 \geq 1/\alpha$, the result is trivial for $|i-j| \leq 1$. If $|i-j| \geq 2$ then $\mathcal{Y}_i$ and $\mathcal{Y}_j$ are orthogonal, so $Y_j U Y_i = -Y_j U^\perp Y_i$. Then with the first result, we know that for a unit vector $y_i \in \mathcal{Y}_i$ we can write $U^\perp y_i = \sum_k n_k^i$ where $n_k^i \in \mathcal{N}_k$ for $k$ even, $n_k'^i \in \mathcal{N}_k'$ for $k$ odd, and $|n_k^i| \leq C_1 \alpha |i-k|$. Then

$$|Y_j U^\perp y_i| = \left| \sum_k Y_j n_k^i \right| \leq |n_j^{i-1}| + |n_j^i| + |n_{j+1}^i| \leq c_\alpha C_1 \alpha |i-j|.$$

So, we now prove the first part. Let $y_i \in \mathcal{Y}_i$. By definition of $U^\perp$, we can write $U^\perp y_i$ as a linear combination of unit vectors $m_j^i \in \mathcal{N}_j$ for $j$ even and $m_j'^i \in \mathcal{N}_j'$ for $j$ odd. Removing some elements from $\mathbb{N} \cap [1, n_0]$ we obtain a set $S$ such that $\{m_s^i\}_{s \in S}$ is linearly independent. Note that by the previous remark, we are simply assuring that $\mathcal{N}_s, \mathcal{N}_s' \neq \emptyset$ for all $s \in S$.

What we want to do is to take advantage of the properties $\mathcal{N}_s'$. Because $N^c \leq U^\perp$, we will isolate the $a_s^i$ for $s$ odd by applying $(1 - N^c)$ to our representation of $U^\perp y_i$. This gives $(1 - N^c) U^\perp y_i = \sum_{s \text{ odd}} a_s^i (1 - N^c) m_s^i$. We want to find relationships between the inner products of the terms $(1 - N^c) m_s^i$ for $s$ odd and use them to obtain control over the $a_s^i$ for $s$ odd. We then extend this control to $s$ even.

We focus on $s, t$ odd. The first statement of Lemma 14.5 implies that $N_s N_t = 0$ if $|a-b| \geq 2$. Consequently $(1 - N^c) m_s^i = (1 - N_{s-1} - N_{s+1}) m_s^i = m_s^i - N_{s-1} m_s^i - N_{s+1} m_s^i \in \mathcal{N}_{[s-1, s+1]}$. So, if $(1 - N^c) m_s^i (1 - N^c) m_t^i \neq 0$ then $\text{dist}([s-1, s+1], [t-1, t+1]) \leq 1$ and since $s, t$ are odd, we have $|s-t| \leq 2$. In the case $s = t$, we have $|(1 - N^c) m_s^i|^2 \geq 1/2 - \eta$ by Equation (24).

For $|s-t| = 2$, we can assume that $t = s + 2$. Then, roughly speaking, the only contribution to the inner product comes from the orthogonality of the ranges of $N_{s+1} N_s'$ and $N_{s+1} N_{s+2}'$ which will give some control by the last statement of Lemma...
In more detail, we have the orthogonal decomposition, \( m^1_s = Y_{s-1}^\prime m^1_s + Y_{s+1}^\prime m^1_s + (1 - Y_{s-1}^\prime - Y_{s+1}^\prime) m^1_s \) from which we can set \( c_s = |Y_{s-1}^\prime m^1_s| \) and \( d_s = |Y_{s+1}^\prime m^1_s| \) so that \( c_s^2 + d_s^2 \leq 1 \). Recall that \( t = s + 2 \) so \( N_{s+1} = N_{t-1} \) and using the fact that the \( Y_\alpha \) commute we see that

\[
((1 - N^e)m^i_s, (1 - N^e)m^i_{s+2}) = |((1 - N^e)m^i_s, m^i_{s+2}) + ((1 - N^e)m^i_s, N^e m^i_{s+2})|
\]

By Proposition 4.2 shows that there are constants \( \tilde{\alpha}_1 \) such that \( |a^{i}_{s}| \leq \tilde{\alpha}_1 \). We will obtain a representation of \( M_{k,l} = \frac{2(1+x)}{1-2\eta}((1 - N^e)m^i_{s_k}, (1 - N^e)U^\perp y_i) \), \( \tilde{a}_i = a^i_{s_i} \), and 

\[
M = \frac{2(1+x)}{1-2\eta}((1 - N^e)m^i_{s_k}, (1 - N^e)m^i_{s_k}) \), \( \tilde{M} = \tilde{a} \). Once we show that \( M \) is invertible, we will obtain a representation of \( \tilde{a} \) as \( M^{-1} \tilde{v} \) and then will use the properties of \( M \) to get control of the \( a^i_s \).

Note that the scaling factor is so that the bounds on the diagonal entries of \( M \) are much clearer. In particular, by construction \( M \) is a tridiagonal self-adjoint matrix and we have the inequalities: \( M_{k,k} \geq \frac{2(1+x)}{1-2\eta}(1/2 - \eta) = 1 + x \geq (c_s^2 + d_s^2) + \) and 

\[
|M_{k,k+1}| \leq \frac{(1+x)(1-\eta)(2-2\eta)}{1-2\eta} d_s c_{s+1} \text{ if } s_{k+1} = s_k + 2 \text{ and } M_{k,k+1} = 0 \text{ otherwise.}
\]

Now, because \( \eta < \chi/4 \), we have that \( x = \frac{\chi}{2-2\chi} > 0 \) so \( M_{k,k} \geq 1 + x \) and \( M_{k,k+1} \leq d_s c_{s+1} \). By Lemma 4.4 \( M - xI \) is positive, so all the eigenvalues of \( M \) are at least \( x \).

Proposition 4.2 shows that there are constants \( \tilde{C} > 0, \tilde{\alpha} \in (0,1) \) such that \( |(M^{-1})_{l,k}| \leq \tilde{C} \tilde{\alpha}^{l-1} \) and these constants only depend on \( x \) and an upper bound on the spectrum of \( M \), which can be taken to be \( 7(1 + x)/(1 - 2\eta) \).

Now, \( N^e \) and \( U^\perp \) commute and \( U^\perp m^i_{s_k} = m^i_{s_k} \) so we see that

\[
U^\perp (1 - N^e)m^i_{s_k} = (1 - N^e)U^\perp m^i_{s_k} = (1 - N^e)m^i_{s_k} \in \mathcal{N}_{[s_k-1,s_k+1]} \subset \mathcal{Y}_{[s_k-1,s_k+1]}
\]

Because \( \tilde{v}_k = \frac{2(1+x)}{1-2\eta}((1 - N^e)m^i_{s_k}U^\perp y_i) \) and \( y_i \in \mathcal{Y}_i \), we see that \( \tilde{v}_k = 0 \) for \( |s_k - i| \geq 3 \). The increasing sequence of odd integers \( s_1, s_2, \ldots \) might have gaps. These gaps cause \( M \) to be a block diagonal matrix, each block being a tridiagonal matrix. Then \( M^{-1} \) also has this same block structure with exponential decay of the entries away from the diagonal. We illustrate this with the example that \( s_1 = 5, s_2 = 7, s_3 = 11, s_4 = \ldots \).
13, $s_5 = 15$ and so $M$ has the following block structure:

$$
\begin{pmatrix}
M_{1,1} & M_{1,2} \\
M_{2,1} & M_{2,2}
\end{pmatrix}
\begin{pmatrix}
M_{3,3} & M_{3,4} \\
M_{4,3} & M_{4,4} & M_{4,5} \\
M_{5,4} & M_{5,5}
\end{pmatrix}.
$$

When bounding $a_k^i$ we can restrict to the block of $M$ in which $s_l$ lies. Let this block be $s_a, \ldots, s_b$. Then for $a \leq k, k' \leq b$, $|s_k - s_{k'}| = 2|k - k'|$. We also have the inequality

$$
|s_l - i| - |s_l - s_k| \leq |s_k - i|.
$$

So,

$$
|a_k^i| \leq \sum_{a \leq k \leq b} |(M^{-1})_{l,k}||\tilde{v}_k| \leq \frac{2(1 + x)}{1 - 2\eta} |y_k| \sum_{a \leq k \leq b} |(M^{-1})_{l,k}| \leq \frac{2(1 + x)}{1 - 2\eta} \tilde{C}|y_k| \sum_{|s_k - i| \leq 2} \tilde{\alpha}|s_i - s_k|/2 \leq \frac{2(1 + x)}{1 - 2\eta} 3\tilde{C}|y_k| \alpha^{-1}\tilde{\alpha}|s_i - i|/2 =: \tilde{\tilde{C}}\alpha^{|s_i - i|}|y_k|.
$$

Now, we can use the smallness of the odd coefficients to deduce the smallness of the even coefficients. If $j$ is even and $|i - j| \geq 2$, because $Y_jy_i = 0$ we see that because the $m_k^i$ are unit vectors,

$$
|a_j^i| = |Y_j(\sum_{s \text{ even}} a_k^i m_s^i)| = |Y_j(y_i - \sum_{s \text{ odd}} a_k^i m_s^i)| = |Y_j(a_j^{i-1} m_j^{i-1} + a_j^{i+1} m_j^{i+1})| \leq \tilde{\tilde{C}}(\alpha^{|i-j|+1} + \alpha^{|i-j|-1}) = \tilde{\tilde{C}}(\alpha + \alpha^{-1})\alpha^{|i-j|}.
$$

Setting $C_1 = \tilde{\tilde{C}}(\alpha + \alpha^{-1})$, we obtain the result. \( \square \)

\section{Properties of $U$ and $W$}

We prove some properties of $U$ and $W$ here that will be used to complete the proof of Lemma \[7.1\]. We begin with the first property that was motivated by Section 12.

This result is stated (albeit with a different exponent of $l_b$) in [1] and the proof is based on that in [1], supplemented by adjustments from [3].

\begin{lemma}
There is an (explicit) constant $C_3$ depending only on $\eta$ such that for $l_b$ large and $u \in U$,

$$
|Au| \geq \sqrt{\frac{1}{C_3 l_b}}|u|.
$$

It suffices that $l_b$ is large enough that $n_bT(l_b) < ((1 - \sqrt{1 - \eta})/6)^2$ and also large enough specified by Lemma \[4.2\].
\end{lemma}
Proof. Recall that $|Au| = (u, \rho u)$. It suffices to show that there is a constant $C < 1$ such that for $w$ in the range of $E_{[0,1/l_b]}(\rho)$, $|Uw| \leq C|w|$. This is because if this is so then using the orthonormal eigendecomposition of $\rho$ expressed in bra-ket notation $\rho = \sum \lambda |v_\lambda\rangle\langle v_\lambda|$ gives

\[
(u, \rho u) \geq \frac{1}{l_b} \sum_{\lambda > 1/l_b} |\langle v_\lambda, u \rangle|^2 = \frac{1}{l_b} |E_{(1/l_b, \infty)}(\rho) u\rangle|^2 = \frac{1}{l_b} (|u|^2 - |E_{[0,1/l_b]}(\rho) u\rangle|^2)
\]

\[
= \frac{1}{l_b} (|u|^2 - \max_{|v|=1} |\langle v, E_{[0,1/l_b]}(\rho) u\rangle|^2) = \frac{1}{l_b} \left(|u|^2 - \max_{E_{[0,1/l_b]}(\rho) v = v\forall |v|=1} |\langle v, u \rangle|^2\right)
\]

\[
\geq \frac{1}{l_b} (|u|^2 - C^2|u|^2) = \frac{1}{l_b} (1 - C^2)|u|^2.
\] (30)

If $w$ is in the range of $E_{[0,1/l_b]}(\rho)$, we have $n_i \in \mathcal{N}_i$, satisfying Item 3 of Lemma 14.5. Because $U^\perp \geq N^\prime$, $N^\prime_i$, we see that $U n_i = 0$ for $i$ even and $U n_i = U(N_i - N^\prime_i)n_i$ for $i$ odd. Now, like in the proof of Lemma 14.12, we will obtain control of the terms $n_i$ for $i$ odd and then extend this control to the rest of terms. The difficulty is that there is no clear way to obtain an approximate decomposition as below of $w \sim w^{\text{even}} + w^{\text{odd}}$ that are approximately (for $l_b$ or $n_b$ large) orthogonal. So, we work around this.

Let $w^{\text{even}} = \sum_{i \text{ even}} n_i$ and $w^{\text{odd}} = \sum_{i \text{ odd}} n_i$, so $w$ is approximately equal to $w^{\text{even}} + w^{\text{odd}}$, as $|w - w^{\text{even}} - w^{\text{odd}}| \leq T(l_b)\sqrt{n_b}|w|$. Also,

\[
|Uw - Uw^{\text{odd}}| = |U(w - \sum_i n_i)| \leq |w - \sum_i n_i| \leq T(l_b)\sqrt{n_b}|w|.
\]

For $s = 1, 3$, let $[s]$ be the set of all natural numbers equivalent to $s$ modulo four. Define $w^1$ and $w^3$ by $w^s = \sum_{i \in [s]} (N_i - N^\prime_i)n_i$. So, $w^s$ is expressed as a series of orthogonal vectors and such that $U(w^1 + w^3) = U(N_i - N^\prime_i)w^{\text{odd}} = Uw^{\text{odd}}$.

Now because $(w^1, w^3) = 0,$

\[
|U^\perp (w^1 + w^3)|^2 = |U^\perp w^1|^2 + |U^\perp w^3|^2 + 2 \text{Re}(U^\perp w^1, w^3) \\
\geq |U^\perp w^1|^2 + |U^\perp w^3|^2 - 2|U^\perp w^1, w^3|^2 \geq |U^\perp w^1|^2 + |U^\perp w^3|^2 - |U^\perp w^1|^2 - |U^\perp w^3|^2.
\]

By Equation (25), for a unit vector $n \in \mathcal{N}_i \cap \mathcal{N}_j$ we know that $|U^\perp n|^2 \geq |N^\prime n|^2 > 1/2 + \eta$. We can apply this to $n = (N_i - N^\prime_i)n_i$. So, if $i, j \in [s]$ are not equal then $|i - j| \geq 4$. Then if $n_i \in \mathcal{N}_i$ then $U^\perp (N_i - N^\prime_i)n_i = (N_{i+1} + N_{i-1})(N_i - N^\prime_i)n_i \in \mathcal{N}_{i+1, i-1}$ and hence $U^\perp (N_i - N^\prime_i)n_i \perp U^\perp (N_j - N^\prime_j)n_j$. This implies that

\[
|U^\perp w^s|^2 = \sum_{i \in [s]} |U^\perp (N_i - N^\prime_i)n_i|^2 = \sum_{i \in [s]} |U^\perp (N_i - N^\prime_i)n_i|^2 \\
\geq (1/2 + \eta) \sum_{i \in [s]} |(N_i - N^\prime_i)n_i|^2 = (1/2 + \eta)|w^s|^2.
\]

Therefore,

\[
0 \leq 2(|U^\perp w^s|^2 - (1/2 + \eta)|w^s|^2) = 2|U^\perp w^s|^2 - |w^s|^2 - 2\eta|w^s|^2 \\
= |U^\perp w^s|^2 - |w^s|^2 - 2\eta|w^s|^2.
\]
Hence we obtain \(|U^\perp(w^1 + w^3)|^2 \geq 2\eta(|w^1|^2 + |w^3|^2) = 2\eta|w^1 + w^3|^2\), so
\[
|Uw| \leq |U(w^1 + w^3)| + T(l_b)\sqrt{\eta}w^3 |w| \leq \sqrt{1 - 2\eta}|w^1 + w^3| + T(l_b)\sqrt{\eta}w
\]
\[
= \sqrt{1 - 2\eta}|w^{\text{odd}}| + T(l_b)\sqrt{\eta}w.
\]

Now, we now will try to give an upper bound of \(|Uw|\) in terms of \(|w|\) based on the cases when \(|N^e w|\) is small and when it is not really that small. When \(|N^e w|/|w|\) is not too small, we can use
\[
|Uw| = |U(1 - N^e)w| \leq |(1 - N^e)w| = \sqrt{|w|^2 - |N^e w|^2}.
\]

When \(|N^e w|/|w|\) is small, we remember the third equation of Item 3 of Lemma 14.5 which gives \(|w^{\text{even}}| \leq \sqrt{|N^e w||w| + T(l_b)\sqrt{\eta}w|w|^2}\) and hence by what we have done:
\[
|Uw| \leq \sqrt{1 - 2\eta}|w^{\text{odd}}| + T(l_b)\sqrt{\eta}w|w| \leq \sqrt{1 - 2\eta}|w - w^{\text{even}}| + (1 + \sqrt{1 - 2\eta})T(l_b)\sqrt{\eta}w
\]
\[
\leq \sqrt{1 - 2\eta}w + \sqrt{1 - 2\eta}\sqrt{|N^e w||w| + T(l_b)\sqrt{\eta}w|w|^2} + 2T(l_b)\sqrt{\eta}w
\]
\[
\leq \sqrt{1 - 2\eta}w + (1 + \sqrt{1 - 2\eta}w + 2\sqrt{\eta}w)T(l_b)|w|,
\]

if we impose \(T(l_b) \leq 1\). Recall that \(\eta < \chi/4 < 1/4\) and consider \(p = \left(\sqrt{\frac{1-\eta}{1-2\eta}} - 1\right)^2\). Then \(p > 0\) and because \(\eta < 3/7, p < 1\). Then for \(|N^e w| \leq p|w|\), the latter estimate gives \(|Uw| \leq \sqrt{1 - \eta}w|w| + 3\sqrt{\eta}T(l_b)|w|\). If \(|N^e w| > p|w|\), then the first estimate gives \(|Uw| \leq \sqrt{1 - p^2}|w|\).

So, picking \(l_b\) large enough that \(3\sqrt{\eta}T(l_b) < (1 - \sqrt{1 - \eta})/2\) and we obtain \(C_3\) using (30) and \(C = \max\left(\frac{1+\sqrt{1-\eta}}{2\sqrt{\eta}}, \sqrt{1-p^2}\right)\).

Note that, as discussed above, this result implies that \(A\) is injective on \(U\). Other consequence are the following inequalities.

**Lemma 15.2.** For \(l_b\) large (determined by Lemma 14.2 and Lemma 15.1), we have:

1. For \(w \in W\), there are \(x_i \in X_i\) such that \(w = \sum_i x_i\) and
\[
|w| \geq \frac{1}{C_5 l_b} \sqrt{\sum_i |x_i|^2}.
\]

2. Let \(x \in X\) and \(y \in Y\) such that \(x = Ay\). Then
\[
|(1-P)x| \leq C_4 \sqrt{\frac{G(l_b)}{l_b}}|y|,
\]

where \(C_4 = C_1 \sqrt{2c_\alpha} \left(\frac{1+\alpha}{1-\alpha}\right)\).

3. Let \(x \in X\) and \(x_i \in X_i\) such that \(x = \sum_i x_i\), then
\[
|(1-P)x| \leq C_4 \sqrt{\frac{G(l_b)}{l_b}} \sqrt{\sum_i |x_i|^2}
\]

44
Proof. For the proof of the first item, note that for \( w \in \mathcal{W} \), there is a (unique) \( u \in \mathcal{U} \) such that \( w = Au \) so

\[
|w| \geq \sqrt{\frac{1}{C_3 l_b}} |u|.
\]

Now, \( \mathcal{U} \subset \mathcal{R} \), so there are orthogonal \( r_i \in \mathcal{R}_i \) (because the spaces themselves are orthogonal) such that \( \sum_i r_i = u \), hence \( |u|^2 = \sum_i |r_i|^2 \). Now, \( w = Au = \sum_i Ar_i \), so if we set \( x_i = Ar_i \in \mathcal{X}_i \), we obtain \( w = \sum_i x_i \). Because \( A \) is an isometry when restricted to each \( \mathcal{R}_i \), \( |x_i| = |r_i| \) hence we obtain the estimate

\[
|w| \geq \sqrt{\frac{1}{C_3 l_b}} \sqrt{\sum_i |r_i|^2} = \sqrt{\frac{1}{C_3 l_b}} \sqrt{\sum_i |x_i|^2}.
\]

We now prove the second statement. Because \( x \in \mathcal{X} \), there is a \( y \in \mathcal{Y}(= \mathcal{R}) \) such that \( x = Ay \). Because we chose \( n_b \) odd, we can express \( y \) as the orthogonal set \( \sum_{i:odd} y_i \) with \( y_i = Y_i y \). We find \( n_j^i \in \mathcal{N}_j \) for \( j \) even (resp. \( \mathcal{N}'_j \) for \( j \) odd) such that \( U^\perp \cdot y_i = \sum_j n_j^i \) and \( |n_j^i| \leq C_1 \alpha^{i-j}[y_i] \). The idea is that because \( \alpha^{i-j} \) has exponential decay, the following estimate is like that of an approximation of the identity.

Recall that \( \rho \) is tridiagonal (with respect to the spaces \( \mathcal{R}_i \)) so \( (n_j^{i_1}, \rho n_j^{i_2}) = 0 \) if \( |j_1 - j_2| \geq 2 \). We now finish the proof of the second statement of the lemma with the following estimate

\[
|(1-P)x|^2 = |(1-P)AU^\perp y|^2 = |A \sum_i U^\perp y_i|^2
\]

\[
= |A \sum_i \sum_j n_j^i|^2 = (\sum_i \sum_j n_j^i, \rho \sum_i \sum_j n_j^i) = \sum_{i,j_1,j_2,|j_2-j_1| \leq 1} (n_j^{i_1}, \rho n_j^{i_2})
\]

\[
\leq \sum_{i,j_1,j_2,|j_2-j_1| \leq 1} (n_j^{i_1}, \rho n_j^{i_2}) \frac{1}{l_b} = \frac{2G(l_b)}{l_b} \sum_{i,j_1,j_2,|j_2-j_1| \leq 1} |n_j^{i_1}|^2
\]

\[
\leq \frac{2c_1 \alpha^{i-j}[y_1]}{l_b} \sum_{i,j_1,j_2,|j_2-j_1| \leq 1} \alpha^{i-j_1}[y_1] \alpha^{i-j_2}[y_2]
\]

\[
= \frac{2c_1 \alpha^{i-j}[y_1]}{l_b} \sum_{j} \left( \sum_i \alpha^{i-j}[y_i] \right)^2 = \frac{2c_1 \alpha^{i-j}[y_1]}{l_b} \sum_{j} \left( \sum_i \alpha^{i-j}[y_i] \right)^2
\]

\[
\leq \frac{2c_1 \alpha^{i-j}[y_1]}{l_b} \sum_{j} \left( \sum_i \alpha^{i-j}[y_i] \right)^2 = \frac{2c_1 \alpha^{i-j}[y_1]}{l_b} \left( \frac{1 + \alpha}{1 - \alpha} \right)^2 |y|^2,
\]

where we have used Minkowski’s inequality (Theorem 1.2.10 in [55]) and

\[
||\alpha^{i-j}||_{L^1} = 1 + 2\alpha \sum_{i \geq 0} \alpha^i = 1 + \frac{2\alpha}{1 - \alpha} = \frac{1 + \alpha}{1 - \alpha}
\]

The last statement of the lemma follows from the second, because we can pick \( r_i \in \mathcal{R}_i \) such that \( Ar_i = x_i \) so then \( |r_i|^2 = |x_i|^2 \). Setting \( y = \sum_i r_i \) gives \( Ay = x \) and \( |y|^2 = \sum_i |r_i|^2 \) so that

\[
|(1-P)x| \leq C_4 \sqrt{\frac{G(l_b)}{l_b}} |y| = C_4 \sqrt{\frac{G(l_b)}{l_b}} \sqrt{\sum_i |r_i|^2} = C_4 \sqrt{\frac{G(l_b)}{l_b}} \sqrt{\sum_i |x_i|^2}.
\]
16 Verifying Items 1, 2, and 3.

This section completes the proof of Lemma 15.1. Note the similarities between the calculations for the spaces $\tilde{X}$ in Section 15.

We first verify Item 1: $\|P^\bot P v_1\| \leq \epsilon_1$. Let $v \in V_1$. We will find a particular $x \in X$ such that $|v - x|$ is small and hence apply the third item of Lemma 15.2 to

$$|(1 - P)v| \leq |(1 - P)(v - x)| + |(1 - P)x|.$$  

Because $\sum_i \tau_i = \sum_i F^0_{\omega(i)}(J)S_1 = S_1$, there is a $x \in \mathbb{C}^d$ such that $v = S_1 x = \sum_i \tau_i x$. Because $S_1$ is an isometry $|v| = |x|$. Then $x_i = \tau_i (1 - Z_i) x \in X_i$ and

$$|v - \sum_{i=0}^{n_{\text{win}}} x_i|^2 = |\sum_{i=0}^{n_{\text{win}}} \tau_i Z_i x|^2 \leq 2 \sum_{i=0}^{n_{\text{win}}} |\tau_i Z_i x|^2 \leq 2 \lambda_{\text{min}}(n_{\text{win}} + 1)|x|^2 = 2 \lambda_{\text{min}}(n_{\text{win}} + 1)|v|^2,$$

where the constant 2 comes from the calculation in equation (1), because the ranges of $\tau_i$ and $\tau_j$ are nonconsecutively orthogonal. Because $x_i = \tau_i (1 - Z_i) x$, we have that

$$\sum_i |x_i|^2 \leq \sum_i |\tau_i x|^2 = \sum_{i, \text{odd}} |\tau_i x|^2 + \sum_{i, \text{even}} |\tau_i x|^2 = |\sum_{i, \text{odd}} F^0_{\omega(i)}(J)v|^2 + |\sum_{i, \text{even}} F^0_{\omega(i)}(J)v|^2 \leq 2|v|^2,$$

because the functions $F^0_{\omega(i)}(t)$ have nonconsecutively disjoint supports so $\|\sum_{i, \text{even}} F^0_{\omega(i)}(J)\|$, $\|\sum_{i, \text{odd}} F^0_{\omega(i)}(J)\| \leq 1$. Then by the third item of Lemma 15.2, we have that

$$|(1 - P)v| \leq |(1 - P) \sum_i x_i| + |v - \sum_i x_i| \leq C_4 \sqrt{\frac{G(l_b)}{l_b}} \sqrt{\sum_i |x_i|^2 + \sqrt{2 \lambda_{\text{min}}(n_{\text{win}} + 1)|v|}}$$

$$\leq \left(C_4 \sqrt{\frac{2G(l_b)}{l_b}} + \sqrt{2 \lambda_{\text{min}}(n_{\text{win}} + 1)}\right)|v|.$$  

(31)

We verify Item 2: $\|(1 - P)JP\| \leq \epsilon_1$. Let $w \in W$. There is a $u \in U \subset Y$ such that $w = Au = AUu$. Write $u$ as the orthogonal sum $\sum_{i, \text{odd}} y_i$, where $y_i = Y_i u$. Then

$$(1 - P)Jw = \sum_i (1 - P)JAU y_i.$$  

Now, we want to use the following two facts. First, each $X_i$ is approximately an eigenspace for $J$ with eigenvalue $\omega(k)$ so each $Y_i = X_i |_{(i-1)b,(i+1)b)}$ (except $i = n_b$, where it is $X_i |_{(i-1)b,n_{\text{win}}} |_{(i-1)b}$ having length less than $4l_b$) is also an approximate eigenspace with eigenvalue $\omega(il_b)$. Second, there is exponential decay for $Y_i U Y_i$. Then we note that each $AU y_i \in W$ so $(1 - P)J(AU y_i) = (1 - P)[J - \omega(il_b)](AU y_i)$. This is the place that we use any special property of the projection $P$, because we then remove it by using $\|1 - P\| \leq 1$. This has a similar feel to the proof of the second item of Lemma 15.2
In more detail, by Lemma 14.12 let \( y_j' = Y_j U y_i \in \mathcal{Y}_j \) for \( j \) odd so \( U y_i = \sum_{j \text{ odd}} y_j' \) and \( |y_j'| \leq C_2 \alpha^{i-j} |y_i| \). Then

\[
|(1-P)w|^2 = \left| \sum_i (1-P)[J - \omega(i b)]Au y_i \right|^2 = \left| (1-P) \sum_{i} \sum_{j \text{ odd}} [J - \omega(i b)]Ay_j' \right|^2
\]

\[
\leq \sum_{i} \sum_{j \text{ odd}} \left| [J - \omega(i b)]Ay_j' \right|^2 = \sum_{i_1,i_2 j_1,j_2 \text{ odd}} \left( [J - \omega(i_1 l b)]Ay_{j_1}^{i_1}, [J - \omega(i_2 l b)]Ay_{j_2}^{i_2} \right).
\]

Note \( Ay_j' \in X_j^{(j-1)l b,(j+1)l b} \) for \( j < n_b \). The following work all applies when \( j = n_b \) except that the upper limits for intervals and sums are both \( n_{\text{win}} \). Because \( X_k \) is a subspace of the range of \( F_{\omega(k)}^{0,\kappa}(J) \), we see that both the \( X_k \) and the \( J(X_k) \) are nonconsecutively orthogonal. Consequently, we continue our calculation as

\[
|(1-P)w|^2 \leq \sum_{i_1,i_2 j_2 \text{ odd}} \sum_{j_1 \text{ odd} |j_2-j_1| \leq 2} \left( [J - \omega(i_1 l b)]Ay_{j_1}^{i_1}, [J - \omega(i_2 l b)]Ay_{j_2}^{i_2} \right). \quad (32)
\]

We now estimate \( \left| [J - \omega(i l b)]Ay_j' \right|^2 \). Because \( y_j' \in \mathcal{Y}_j = R_{(j-1)l b,(j+1)l b} \), we write it as an orthogonal sum \( \sum_{k=(j-1)l b}^{(j+1)l b} r_k \) for \( r_k \in R_k \). We then obtain

\[
\left| [J - \omega(i l b)]Ay_j' \right|^2 = \left| [J - \omega(i l b)] \sum_{k} Ar_k \right|^2 = \sum_{k_1,k_2} \left( [J - \omega(i l b)]Ar_{k_1}, [J - \omega(i l b)]Ar_{k_2} \right).
\]

Note that \( J - \omega(k) \) restricted to \( X_k \) has norm bounded by \( \kappa \). Now, \( R_k \subset \mathcal{Y}_j \) for some \( j \) so

\[
\left| [J - \omega(i l b)]Ar_k \right| \leq \left| [J - \omega(k)]Ar_k + |\omega(k) - \omega(i l b)||Ar_k| + |\omega(i l b) - \omega(j l b)||Ar_k| \right|
\]

\[
\leq 2\kappa(1 + (1 + |i-j|)l b)|Ar_k| = 2\kappa(1 + (1 + |i-j|)l b)|r_k|.
\]

Let \( C_{i,j} := 2\kappa(1 + (2 + |i-j|)l b) \) to account for the cases \( j < n_b \) and \( j = n_b \). So, keeping in mind that the range of \( k \), (we can define \( r_k = 0 \) outside this range) we see that

\[
\left| [J - \omega(i l b)]Ay_j' \right|^2 = \sum_{k_1} \sum_{|k_1-k_2| \leq 1} \left( [J - \omega(i l b)]Ar_{k_1}, [J - \omega(i l b)]Ar_{k_2} \right)
\]

\[
\leq C_{i,j}^2 \sum_{k_1} \sum_{|k_1-k_2| \leq 1} |r_{k_1}| |r_{k_2}| = C_{i,j}^2 \sum_{k} |r_k|^2 \sum_{|k+k| < 1} |r_{k+\sigma}|
\]

\[
\leq C_{i,j}^2 \sum_{|k+k| < 1} \left( \sum_{k} |r_k|^2 \right)^{1/2} \left( \sum_{|k+k| < 1} |r_{k+\sigma}|^2 \right)^{1/2}
\]

\[
\leq 3C_{i,j}^2 \sum_k |r_k|^2 = 3C_{i,j}^2 |y_j'|^2.
\]
Using $2\sqrt{3}\kappa(1 + (2 + |i - j|)b_0) \leq 2\sqrt{3}(\kappa b_0)(3 + |i - j|) =: K(3 + |i - j|)$ and $C_\alpha = \max_x(|x| + 3)\alpha|x|/2$, we insert our above calculations into Equation (32) to get

$$|(1 - P)Jw|^2 \leq K^2 \sum_{i_1, i_2, j_2 \text{ odd}} \sum_{j_1 \text{ odd}, |j_2 - j_1| \leq 2} (3 + |i_1 - j_1|)(3 + |i_2 - j_2|)|y_{i_1}||y_{i_2}|.$$

$$\leq (C_2 K)^2 \sum_{i_1, i_2, j_2 \text{ odd}} \sum_{j_1 \text{ odd}, |j_2 - j_1| \leq 2} (3 + |i_1 - j_1|)(3 + |i_2 - j_2|)|y_{i_1}|\alpha^{i_1 - j_1}|y_{i_2}|\alpha^{i_2 - j_2}|$$

$$\leq c_\alpha(C_\alpha C_2 K)^2 \sum_{j \text{ odd}} \sum_{i_1, i_2} |y_{i_1}|\alpha^{i_1 - j/2}|y_{i_2}|\alpha^{i_2 - j/2} = c_\alpha(C_\alpha C_2 K)^2 \sum_{j \text{ odd}} \left(\sum_i |y_i|\alpha^{i - j/2}\right)^2$$

$$\leq c_\alpha(C_\alpha C_2 K)^2 \left(\frac{2 + \alpha}{2 - \alpha}\right)^2 \sum_k |y_k|^2 = c_\alpha K^2 C_\alpha^2 C_2^2 \left(\frac{2 + \alpha}{2 - \alpha}\right)^2 |u|^2 =: (K_\alpha n b_0 |u|^2).$$

Because $u \in U$ with $w = Au$, Lemma 15.1 shows that

$$|(1 - P)Jw| \leq K_\alpha \kappa n b_0 \sqrt{C_3 b_0 |w|} = \text{Const.} n b_0^{3/2} |w|.$$

Now, for the third item, we use the alternate characterization in Remark 7.3. So, let $w \in W$ and we will bound $P_{V_L}(w)$. Now, by Lemma 15.2 we can write $w = \sum x_i$, $x_i \in X_i$ such that $\sqrt{\sum_i |x_i|^2} \leq \sqrt{C_3 b_0 |w|}$. (33)

We will bound each $P_{V_L}(x_i)$ using the Lieb-Robinson estimates. Let $\hat{B}$ be a position operator on $B$ being $jI$ on $V_j$. Then $J$ is tridiagonal with respect to these blocks so it satisfies the conditions of Corollary 6.7 with $\Delta = 2$. So,

$$\|P_{V_L} F_{\omega(i)}^0(J) P_{V_1}\| \leq \int_{|k| \geq \frac{L - 1}{2\pi}} |\hat{F}_{0,k}(k)|dk + \|\hat{F}_{0,k}\|_{L^1(\mathbb{R})}e^{-\frac{|k|}{4\pi}}.$$ 

By equations (17) and (20) and the definition of $\kappa$,

$$\|P_{V_L} F_{\omega(i)}^{0,k}(J) P_{V_1}\| \leq \int_{|k| \geq \frac{L - 1}{2\pi}} |\hat{F}_{0,k}^{0,1}(k)|dk + \|\hat{F}_{0,k}^{0,1}\|_{L^1(\mathbb{R})}e^{-\frac{|k|}{4\pi}} = S(L),$$

as defined in Section 3. Now, since $x_i \in X_i$, we have that there are $x_i \in \mathbb{C}^d$ such that $x_i = F_{\omega(i)}^{0,k}(J) S_1(1 - Z_i)x_i$. We can pick $x_i$ in the kernel of $Z_i$ so that $x_i = F_{\omega(i)}^{0,k}(J) S_1 x_i$, with $S_1 x_i \in V_1$ and $|x_i|^2 = (\tau^* \tau x_i, x_i) \geq \lambda_{\min} |x_i|^2 = \lambda_{\min} |S_1 x_i|^2$. So,

$$|P_{V_L} x_i| = |P_{V_L} F_{\omega(i)}^{0,k}(J) S_1 x_i| \leq S(L) |S_1 x_i| \leq \frac{S(L)}{\lambda_{\min}^{1/2}} |x_i|.$$ 

By Equation (33), we have

$$|P_{V_L}(w)| \leq \sqrt{n_{\text{win}} + 1} \left(\sum_{i=0}^{n_{\text{win}}} |P_{V_L}(x_i)|^2\right)^{1/2} \leq S(L) \sqrt{\frac{C_3(n_{\text{win}} + 1)b_0}{\lambda_{\min}}} |w|. $$

48
So that we have gotten our estimates, we write \( n_b \sim L^{\beta_0}, n_{\text{win}} \sim L^{\beta_1}/F(L), \lambda_{\text{min}} \sim 1/(L^{\beta_2}(n_{\text{win}} + 1)) \sim F(L)/L^{\beta_1+\beta_2} \), for \( \beta_0, \beta_1, \beta_2 > 0 \), and in the definition of \( n_{\text{win}} \) we had \( \beta_1 \leq 1 \). Recall that \( l_b \sim n_{\text{win}}/n_b \sim L^{\beta_1-\beta_0}/F(L), \kappa \sim F(L)/L^{\beta_1} \).

We then get

\[
\epsilon_3 \leq \sqrt{G(L^{\beta_1-\beta_0}/F(L))F(L)L^{\beta_0/2-\beta_1/2} + L^{-\beta_2/2}},
\]

\[
\epsilon_4 \leq \text{Const.} \frac{F(L)}{L^{\beta_1}} \left( \frac{L^{\beta_1-\beta_0}}{F(L)} \right)^{3/2} \sim \text{Const.} \frac{L^{\beta_1/2-3\beta_0/2}}{F(L)^{1/2}},
\]

and

\[
\epsilon_5 \leq \text{Const.} S(L) \sqrt{\frac{(n_{\text{win}} + 1)l_b}{\lambda_{\text{min}}}} \sim \text{Const.} \frac{S(L)}{F(L)^{3/2}} L^{3\beta_1/2+\beta_2/2-\beta_0/2}.
\]

So, if \( \epsilon_3 \) and \( \epsilon_4 \) have similar rates, because we are assuming that \( G(l) \) and \( F(L) \) grow slower than any polynomial, we get \( \beta_0 - \beta_1 = \beta_1 - 3\beta_0 = -\beta_2 \), hence \( \beta_0 = \beta_1/2 \). This gives a rate of \( L^{-\beta_1/4} \). However, since \( \beta_1 \leq 1 \), the best that this gives us is \( L^{-1/4} \), which is why we pick \( \beta_1 = 1 \). We pick \( \beta_2 = 1/2 \) and \( \beta_0 = 1/2 \). We get \( \gamma_2 = 1/4 \).

This ends the proof of Lemma 7.1.

17 Constructive Aspects

A main issue with finding nearby commuting matrices numerically given some pair of almost commuting matrices is that we would need to find matrices that exactly commute. Even if our algorithm generates commuting matrices in a “pure” mathematical sense, we might only ever generate nearly commuting matrices (although they might have a much small commutator than the original pair of almost commuting matrices). Moreover, we might not be able to distinguish numerically between an algorithm that actually solves this problem and one that does not.

Also, there is also not an obvious way of solving this problem in a “brute force” manner. However, if we are guaranteed that a matrix/matrices satisfying certain properties exists/exist then we might be satisfied with an approximation of a solution within a certain error. In particular, we formulate Lin’s theorem and the non-constructive part of Hastings’ construction in terms of inequalities that determine an open set of some parameterizable set of matrices. We first begin with discussing a pair of nearly diagonal matrices.

We first begin with discussing parameterizations of unitaries and subspaces. The introduction to [42] includes applications of parameterizations of unitaries and also some references to other parameterizations. [42] includes parameterizations of unitary matrices and projections on \( \mathbb{C}^n \). Continuing that approach, Theorem 2 in [43] gives parameterizations\(^\text{6}\) of all special unitary matrices \( SU(n) \). [44] contains a representation of \( U(\lambda) \in SU(n) \) in terms of the “parameterization matrix” \( \lambda \in M_n(\mathbb{R}) \) as products of matrices of the form \( e^{iA_{i,j}} \) where \( A \) is self-adjoint. The parameterization is for \( \lambda \in \Lambda := [0, \pi/2]^n \times [0, \pi]^n \times [0, 2\pi]^n \). This implies that

\[
\|U(\lambda) - U(\mu)\| \leq \sum_{i,j} |\lambda_{i,j} - \mu_{i,j}|.
\]

\(^6\)There are MATLAB functions written for these parameterizations of \( U(n) \) and \( SU(n) \) on Mathworks’ website ([44]).
Define
\[ \epsilon(A, B, U) = \max(\|UAU^* - \text{diag}(UAU^*)\|, \|UBU^* - \text{diag}(UBU^*)\|) \]
and
\[ \epsilon(A, B) = \min_{U \in SU(n)} \epsilon(A, B, U). \]
Then the inequality
\[ \| \text{diag}(B) \| = \left\| \sum_{k=1}^{n} |e_k\rangle B \langle e_k| \right\| \leq n \|B\| \]
suggests that finding a minimizer for \( \epsilon(A, B) \) becomes harder to calculate as \( n \) grows.

Specifically, if \( U_0 \) is a minimizer in the definition of \( \epsilon(A, B) \) then for any \( U \in SU(n) \),
\[ |\epsilon(A, B, U_0) - \epsilon(A, B, U)| \leq 2(1 + n)\|U_0 - U\|. \]
This implies that if we choose an explicit \( \frac{\epsilon}{2(1 + n)} \)-dense subset \( \Lambda_0 \) of \( \Lambda \) with respect to the \( \ell^1 \) norm, then we are guaranteed to have a \( \lambda_0 \in \Lambda_0 \) such that \( |\epsilon(A, B, U_0) - \epsilon(A, B, U(\lambda_0))| \leq \epsilon \) by just searching through all elements of \( \Lambda_0 \).

This shows that, in principle, we can find a close commuting pair \( A', B' \) near \( A, B \) according to the following reformulation of Lin’s theorem.

\((Q'')\): For \( A, B \) self-adjoint contractions with small commutator, there is a unitary matrix \( U \) such that \( UAU^*, UBU^* \) are (jointly) nearly diagonal.

**Proposition 17.1.** Lin’s theorem and \((Q'')\) are equivalent.

**Proof.** That is, suppose \( A, B \) have nearby self-adjoint commuting matrices \( A', B' \). Then pick \( U \) to be a unitary matrix such that \( UA'U^*, U'B'^* \) are diagonal. Then \( \|UAU^* - U'A'U'^*\| = \|A - A'\|, \|UBU^* - U'B'^*\| = \|B - B'\| \).

If we have a \( U \) such that \( UAU^*, UBU^* \) are nearly diagonal, then there are diagonal \( D_A, D_B \) such that \( U^*D_AU, U^*D_BU \) commute and \( \|U^*D_AU - A\|, \|U^*D_BU - B\| \) are small.

The following is a “constructive bootstrap” which provides a brute force algorithm for giving a projection \( P \) with certain constraints along the lines of \((Q')\), \((Q'')\) or Lemmas \( \ref{lem:12.1} \) \( \ref{lem:7.4} \) or \( \ref{lem:12.1} \). These results show that if there exists a projection satisfying certain inequalities and then using a parameterization, we can demonstrate a nearby projection that also satisfies slightly weaker inequalities.

**Lemma 17.2.** For any \( \epsilon \in (0, 1) \), there is a \( \delta > 0 \) such that if \( A, B \) are self-adjoint matrices with \( \|A\|, \|B\| \leq 1 \) and \( \|[A, B]\| \leq \delta \) then we can construct a projection \( P \) such that \( \|[P, B]\| \leq \epsilon \) and \( E_{[-1, -1/2]}(A) \leq P \leq 1 - E_{[1/2, 1]}(A) \).

**Proof.** Apply Lemma \( \ref{lem:12.1} \) with \( \epsilon/2 \) instead of \( \epsilon \). This means that the \( \delta > 0 \) (which is undetermined) for this lemma will be less than the \( \delta \) gotten by applying Lemma \( \ref{lem:12.1} \) without being able to construct the projection.

Consider the set \( \mathfrak{P} \) of projections in \( M_n(\mathbb{C}) \) such that \( E_{[-1, -1/2]}(A) \leq P \leq 1 - E_{[1/2, 1]}(A) \). (Note that this set depends only on \( A \).) What Lemma \( \ref{lem:12.1} \) states is that
\[ M(\delta) = \max_{A, B \in B(M_n(\mathbb{C}), \|[A, B]\| < \delta)} \min_{P \in \mathfrak{P}} \|[P, B]\| < \epsilon/2. \]
An algorithm for finding a $P$ in $\mathcal{P}$ such that $\|P, B\| \leq \epsilon$ is to find projections $P_1, \ldots, P_m$ in $\mathcal{P}$ such that $B_{P_k}(\epsilon/4) = \{P \in \mathcal{P} : \|P - P_k\| < \epsilon/4\}$ cover $\mathcal{P}$. Once this is done, we note that by searching through the $P_k$, we can find one such that $\|P_k - P\| < \epsilon/4$ and hence $\|[P_k, B]\| \leq 2\|P_k - P\| + \|[P, B]\| \leq \epsilon$.

The projections $P_1, \ldots, P_m$ can be found as follows. Let $u_1, \ldots, u_r; u_{r+1}, \ldots, u_{s-1};$ and $u_s, \ldots, u_n$ be orthonormal bases for the ranges of $E_{[1, -1/2]}(A), E_{(-1/2, 1/2)}(A),$ and $E_{[1/2, 1]}(A)$, respectively. Let $U$ be the unitary matrix with columns $u_i$.

Let $\mathcal{P}'$ be the set of projections in $M_{n'}(\mathbb{C})$, where $n' = s - r - 1$. This set can be parameterized according to $\mathcal{P}$, where the parameterization from $\mathcal{P}$ of projections of dimension $m$ is gotten by forming the matrix $S_m \in M_{n'}$ whose columns are $e_1, \ldots, e_m$ and multiplying this on the left by parameterized unitary matrices $U_{CS}$, expressed as the product of $2m(n' - m)$ many matrices of the form $e^{iA\lambda_{r,s}}$ where $A$ is self-adjoint. These are the same unitaries as before, except with a restriction on the parameterization matrix, so as to effectively give $\lambda \in [0, 2\pi]^{\binom{n'}{2}} \times [0, \pi/2]^{m(n' - m)}$. As before, we would need a subset of this parameterization space that is $\epsilon/4\sqrt{n'}$-dense with respect to the $\ell^1$ norm.

We do this for all dimensions $m \leq n'$ to obtain finitely many projections $P_k'$. Then set

$$P_k = U \begin{pmatrix} I_r & P_k' \\ P_k' & I_{n-s} \end{pmatrix} U^*.$$

We then let $P$ be a minimizer of

$$\min_k \|[P_k, B]\| \leq \epsilon.$$

This concludes the construction. \qed

18 Almost commuting Hermitian and Normal matrices

Continuing the discussion in Section 2, Davidson in [5] also provided an absorption result which is a weakened version of Lin’s theorem and also different formulations of Lin’s theorem which are discussed in Section 3. The absorption result is [5]’s Theorem 0.1 which states that for $A, B \in M_n(\mathbb{C})$ self-adjoint, there exist commuting self-adjoint $C, D \in M_n(\mathbb{C})$ with $\|C\| \leq \|A\|, \|D\| \leq \|B\|$ and commuting $A', B' \in M_{2n}(\mathbb{C})$ such that

$$\|A \oplus C - A'\|, \|B \oplus D - B'\| \leq 25\|[A, B]\|^{1/2}.$$

In 1988, Choi, using the signature of matrices as an invariant, provided a result similar to that of [5] of an almost commuting self-adjoint diagonal matrix and weighted shift operator where there are no nearby commuting matrices at all.

Earlier in 1983, Voiculescu provided the elementary counter-example for Lin’s the-
orem for unitary matrices (we take the form used in [13]):

\[
U_n = \begin{pmatrix}
\omega_n & \omega_n^2 & \cdots & \\
\omega_n & \omega_n & \cdots & \\
& & \ddots & \\
& & & 1
\end{pmatrix}, \\
V_n = \begin{pmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
& 1 & 0 \\
& & \ddots
\end{pmatrix},
\]

where \(\omega_n = e^{2\pi i/n}\) is the principal \(n\)th root of unity. Using an argument involving spectral projections and a result of Halmos concerning the unilateral shift, he showed that the variant of Lin's theorem where one requires that the contractions be unitary, rather than Hermitian, is false. This example, like many of the other explicit counterexamples, involves a diagonal and weighted shift matrices. Moreover, after his proof of the result, Voiculescu framed the problem of Lin's theorem in terms of liftings of a *-homomorphism \(C(X) \to \mathcal{A}/\mathcal{J}\) to a map \(C(X) \to \mathcal{A}\) and hypothesized (correctly) that the second homology of the two-sphere \(S^2\) might have played a role in his counter-example (which is trivial for the 2-torus \(T^2 = S^1 \times S^1\)) hence “it seems rather improbable that these examples will have direct bearing on the problem of self-adjoint operators.”

Taking up the ideas of Voiculescu concerning liftings, Loring in 1988 in [12] investigated almost commuting unitary matrices using \(K\)-theory. Loring proved a counterexample similar to that of Davidson and conjectured that if the \(K\)-theoretic obstruction vanished then there is a lifting. As an example, he provided an illustrative absorption result similar to that of [5]'s for Voiculescu's unitaries where taking the direct sum with some unitary matrices \(C,D\) removes the \(K\)-theoretic obstruction so that there are nearby commuting unitary matrices. In particular, he proved that there are commuting unitaries \(U,V\) such that \(\|U_{n2} \oplus U_{n2} - U\|, \|V_{n2} \oplus V_{n2}^* - V\| \leq 4\pi/n\).

Then in 1989, Exel and Loring in [13] provided an elementary winding number argument showing that there are no commuting matrices at all that are nearby Voiculescu’s unitaries. This winding number involved taking straight line paths \(U(t), V(t)\) from Voiculescu’s unitaries \(U,V\) to any given commuting matrices \(U',V'\) and calculating for each \(t\), the path sending \(r \in [0,1]\) to

\[
\gamma_t(r) = \det((1-r)U(t)V(t) + rV(t)U(t)).
\]

Using that \([U(1), V(1)] = [U', V'] = 0\), that \(\|[U, V]\| = |1 - \omega_n|\), and elementary properties about winding numbers in the punctured plane, Exel and Loring obtain their result with an explicit bound on how close \(U',V'\) can be to \(U,V\).

Later in [14] in 1991, Exel and Loring showed that a \(K\)-theoretic invariant and the winding number invariant discussed above are the same (when \([U,V]\) is small). The \(K\)-theoretic invariant is similar to Choi’s invariant in that it is the signature of a matrix formed from \(U\) and \(V\).

Section 1 of [9] formulates various almost commuting - nearly commuting problems under certain “geometric” restrictions. These examples include the two that are directly related to the primary reformulations of Lin’s theorem that was not discuss:

1. The geometry of a square, which is \(\{(x_1, x_2) \in \mathbb{R}^2 : |x_1|, |x_2| \leq 1\}\), giving rise to almost commuting Hermitian \(H_1, H_2\) with \(\|H_1\|, \|H_2\| \leq 1\), and
2. The geometry of the disk, which is \( \{ z \in \mathbb{C} : |z| \leq 1 \} \), giving rise to an almost normal \( N \) with \( \|N\| \leq 1 \).

Some other examples (which we will discuss below) are that of almost commuting Hermitian and unitary matrices (which is related to the geometry of a cylinder/anulus) and two almost commuting unitaries (the geometry of the torus). The latter does not always have nearby commuting matrices, but \([11]\) showed that if both unitaries have a spectral gap then one obtains nearby commuting unitaries. The proof given there involves using a matrix logarithm to reduce the unitary matrices with a spectral gap into Hermitian matrices. Similarly, an almost commuting Hermitian \( A \) and unitary \( B \) are nearly commuting and the proof in \([9]\) takes \( N = AB \) as an “almost polar decomposition” of the almost normal matrix \( N \).

Both of the geometries discussed above (and more general geometries) are addressed by Enders and Shulman in \([30]\) providing a dimensional and cohomology criterion for liftings in the spirit of Voiculescu’s comment in \([28]\). In this section we discuss an approach to a few of these types of results following the ideas presented in previous sections.

We make some remarks on the continuity of \( f : \mathbb{R} \to \mathbb{C} \) or \( f : \mathbb{C} \to \mathbb{C} \) as a function on normal matrices. A continuous function \( f \) of a complex variable is operator continuous (see Proposition II.2.3.3 of \([54]\)). It is important to note that the operator modulus of continuity of \( f \) is bounded below by the modulus of continuity of \( f \) as a function of a real/complex variable. The proof in \([54]\) shows that the operator modulus of continuity can be chosen independent of the \( C^*\)-algebra. See \([35]\) for a more detailed analysis of the operator modulus of continuity. As seen in a remark before Lemma 5.2 of \([35]\), which references proofs given in Section 10 of \([36]\), if \( f \) is a continuous function, then its operator modulus of continuity is equivalent to its “commutator modulus of continuity”. That is,

\[
\Omega_{f,\delta}(\delta) = \sup \{ \| f(A) - f(B) \| : A, B \text{ normal}; \sigma(A), \sigma(B) \subset \mathfrak{F}; \| A - B \| \leq \delta \}
\]

is equivalent to

\[
\Omega_{f,\delta}^\prime(\delta) = \{ \| [f(A), R] \| : A \text{ normal}, \sigma(A) \subset \mathfrak{F}, R \text{ self-adjoint}, \| R \| \leq 1, \| [A, R] \| \leq \delta \}.
\]

For a summary of Operator Lipschitz functions and the norm \( \| - \|_{OL(\mathbb{R})} \) see \([34]\). Note that \( \| f \|_{OL(\mathbb{R})} \leq C_f \), which can be seen by using the Fourier representation of \( f \) and \( \| e^{ikx} \|_{OL(\mathbb{R})} = |k| \) from \([34]\). Although \( f \) being a Lipschitz function of a real/complex variable does not ensure that \( f \) is Operator Lipschitz, its operator Lipschitz constant is at least its Lipschitz constant. \([34]\) discusses how \( f \) being a Hölder continuous function of a real variable guarantees that it is Operator Hölder continuous with an equivalent (depending on the Hölder exponent) norm.

We present a naïve non-proof that given two commuting self-adjoint contractions \( A_1, A_2 \) that almost commute with the self-adjoint contraction \( B \), there exist commuting self-adjoint \( A'_1, A'_2, B' \) that are close to \( A_1, A_2, B \), respectively. The idea is that because \( [A_1, A_2] = 0 \), we can simultaneously diagonalize these matrices and hence there exists a self-adjoint matrix \( A \) whose eigenspaces are subspaces of the eigenspaces of \( A_1 \) and the eigenspaces of \( A_2 \). Therefore, we can write \( A_1 = f_1(A), A_2 = f_2(A) \) for some continuous functions \( f_1, f_2 : \mathbb{R} \to \mathbb{R} \). We hope that \( [A, B] \) is small so that there are commuting \( A', B' \) near \( A \) and \( B \), respectively and that if we define \( f_1(A') = A'_1, f_2(A') = A'_2 \) then
A_1' is near A_1 and A_2' is near A_2. It is a given that A_1', A_2', B' all commute, but the other “hopes” are not guaranteed to be realizable. The first issue is the commutator of [A, B] might not be small. The second issue is that we might not have control of how close A_1' = f_i(A') is to A_i if the functions f_i depend on the original matrix A.

Exploring an example more in line with what we will be doing, suppose that the spectrum of a normal matrix N is a subset of the graph \( \{(x, f(x)) : x \in [-1, 1]\} \), where we identify \( \mathbb{R}^2 \) with \( \mathbb{C} \). If f is a continuous function, then set A_1 = (N + N^*)/2, A_2 = (N − N^*)/2\iota. There are commuting Hermitian A_1', B'. Because A_2 = f(A_1), setting A_2' = f(A_1') makes N' = A_1' + iA_2' normal and close to N because \( \|A_2 - A_2'\| = \|f(A_1) - f(A_1')\| \). We then have that N, B are nearly commuting.

What we mean by “a nice function” is that either we have one fixed continuous function f or we have a family operator equicontinuous functions \( \mathcal{F} \) from which we pick f. For instance, we might want to choose a sequence of functions \( f_n \) that are uniformly operator Lipschitz.

To illustrate how the above naive construction can fail, consider N to have spectrum in \([0,1]^2\). We can perturb N (very) slightly so the spectrum is in \( \{(\frac{k}{n}, \frac{j}{n}) : -n \leq k, j \leq n\} \) for some small \( \epsilon > 0 \). Then an interpolating function f will display behavior like a saw-tooth function like with slope at least \( \epsilon \) and hence has growing operator modulus of continuity. A notable example of Davidson in [5] is of a weighted shift operator in \( M_n(\mathbb{C}) \) (where \( n - 1 \) is a square) that is almost normal and almost commutes with a diagonal Hermitian matrix. Following his use of [31], one sees that we get a normal matrix that is unitarily equivalent to \( \bigoplus_{k=0}^{M} c_k S_d_k \), where \( S_m \) is a cyclic permutation matrix of length \( M \); \( M = 2 \left\lfloor \frac{3n^2-2}{2} \right\rfloor \); \( n \geq 16^4 + 2 = 65538 \); \( d_k \geq M/2 \); and \( c_k = \frac{M-k}{M} \). This normal matrix has spectrum in the union of concentric circles of radii \( c_k \) centered at the origin. The limit points of its spectrum as \( n \to \infty \) is the entire unit disk. [5] proved that this normal matrix and a Hermitian matrix do not have nearby commuting matrices.

A way that the above construction works is if the spectrum of N lies within a simple curve \( \Gamma \) with distinct endpoints such that there is a closed (and bounded) interval \( I = [-a, a] \) and continuous functions \( f : I \to \Gamma \subset \mathbb{C} \) and \( g : \mathbb{C} \to I \) such that \( f \) is a bijection and \( g \circ f = \text{id}_I \). Then if N almost commutes with B, \( A = g(N) \) almost commutes with B as discussed above. Then nearby A, B there are commuting Hermitian A', B'. Scaling A' by a number slightly less than 1 if necessary as in [2], we can assume that the spectrum of A is in I. Setting N' = f(A'), we see that N' commutes with B' and

\[
\|N' - N\| = \|f(A') - f(A)\|
\]

is small when \( \|[N, B]\| \) is small. If \( \Gamma \) is the disjoint union of such curves, then one can proceed in a similar way by forming block matrices gotten by collecting the spectrum lying in different curves.

We will explore how a type of one-dimension spectrum of a normal matrix N can guarantee that if N and a Hermitian matrix A almost commute, then there are nearby commuting normal and Hermitian matrices. We first approach the situation of an almost commuting Hermitian contraction A and unitary U.

**Proposition 18.1.** Let A be a Hermitian contraction and U unitary. Then if A, U are almost commuting, they are nearly commuting (with the same asymptotic rates as in Theorem 1.)
Proof. Let $\delta = \|[A, U]\|$. Picking $\Delta = \delta^{70}$, we apply Corollary 6.4 to obtain a finite range (of distance $\Delta$ with respect to the normal $U$) $H$ such that $\|[A - H]\| \leq \text{Const.} \delta^{1 - 70}$ and $\|[H, N]\| \leq \text{Const.} \delta^{70}$.

Recall that in Section 7 where we have $B$ self-adjoint, we partitioned its spectrum into intervals much larger than $\Delta$. We used the tridiagonal nature of $H$ by grouping together eigenvalues of $B$ so as to apply Lemma 7.4 to a “pinching” of $H$ by certain projections to obtain the subspaces $\mathcal{W}$.

So, we do the same here: we partition the spectrum of $U$ into arcs $I_i$ by forming $n_{\text{win}} \sim \Delta^{-\gamma_1}/2\pi$ many intervals, where $0 < \gamma_1 < 1$. This gives the subspaces $\mathcal{B}_i$ on which $H$ restricted is block tridiagonal with many blocks and on which $U$ takes on potentially many eigenvalues, but all in the interval $I_i$. For $H$ projected onto $\mathcal{B}_i$, we then obtain the almost invariant subspaces $\mathcal{W}_i$ satisfying the properties specified in Lemma 7.4 including $\mathcal{B}_i = \mathcal{W}_i^{\perp} \oplus \mathcal{W}_{i+1} \subset \mathcal{B}_i \oplus \mathcal{B}_{i+1}$. We then have that $\mathcal{B}_i$ is almost invariant for $H$. This construction is essentially identical, with the key distinction that although there are “first” and “last” subspaces, we index the spaces $\mathcal{B}_i$ cyclically and must (at least for our construction to work) define $\mathcal{B}_{n_{\text{cut}}} = \mathcal{W}_{n_{\text{cut}}}^{\perp} \oplus \mathcal{W}_{n_{\text{cut}}+1} = \mathcal{W}_{n_{\text{cut}}}^{\perp} \oplus \mathcal{W}_1 = \tilde{\mathcal{B}}_0$.

Let $U'$ be a multiple of the identity on $\tilde{\mathcal{B}}_i$, with the multiple being the center of the arc $I_i$. We project $H$ onto the $\tilde{\mathcal{B}}_i$ to obtain $H'$ and we conclude. \hfill \Box

Note that we then obtain a stronger form of Osborne’s result in [11] only requiring one matrix to have a spectral gap. We only need to transform one unitary into a Hermitian matrix, but we use a modification of a fractional linear transformation instead.

**Proposition 18.2.** Fix $\theta \in (0, 2\pi)$. Then for unitaries $U, V$ where $V$ has a spectral gap of angular radius $\theta$, there are nearby commuting unitaries $U', V'$. If $\epsilon = \epsilon(\delta)$ derived from Lin’s theorem for a unitary and self-adjoint pair, then

$$\max(||U - U'||, ||V - V'||) \leq 2\epsilon \left( \frac{\|[U, V]\|}{1 - \cos \theta} \right).$$

**Proof.** Just as in [11], we can multiply $V$ by a phase so that the spectral gap is centered at 1. Set $f_\theta(x) = \frac{x + 1}{x - 1}$ and $g_\theta(z) = i\frac{1 + z}{1 - z}$. A simple calculation (or a simple comparison to fractional linear transformations in [50]) shows that $f_\theta$ maps $\mathbb{R}$ to the unit circle and $g_\theta$ maps the unit circle to $\mathbb{R}$ as its inverse. Writing $\alpha = e^{i\varphi}$, one can see that

$$g_\theta(e^{i\varphi}) = -\frac{\sin \varphi}{1 - \cos \varphi}.$$

We set $W = g_\theta(V)$. We then have $\|W\| \leq 1$ and using $\frac{x + 1}{x - 1} = 1 + \frac{2}{x - 1}$ we see that

$$\|[U, W]\| = 2\|[U, (V - 1)^{-1}]\| \leq 2\|(V - 1)^{-1}\| ||(V - 1)U(V - 1)^{-1} - U||$$

$$\leq 2\|(V - 1)^{-1}\|^2 \|[U, V - 1]\| \leq \frac{1}{1 - \cos \theta} \|[U, V]\|.$$

We then can find nearby commuting $U', W'$ where $U'$ is unitary and $W'$ is self-adjoint. Then $V' = f_\theta(W')$ is unitary and we have

$$\|V - V'\| \leq \|(W - i)(W + i)^{-1} - (W' - i)(W' + i)^{-1}\|$$

$$= 2\|(W + i)^{-1} - (W' + i)^{-1}\| \leq 2\|W - W'\|.$$
by the resolvent identity calculation in Example 1 of [34]. Then we have \( \| V - V' \| \leq 2\epsilon \left( \frac{\delta}{1 - \cos \theta} \right) \).

Now we can attempt to do the same sort of construction for a normal matrix \( N \) whose spectrum belongs to some set \( \Gamma \). We define the following condition on \( \Gamma \subset \mathbb{C} \), but note that it is not necessarily optimal.

\(*) \Gamma \text{ will be the union of finitely many simple curves } \Gamma^k \text{ which include their two endpoints. Let } \{z_1, \ldots, z_n\} \text{ be the endpoints of the } \Gamma^k. \text{ The curves } \Gamma^k \text{ are not permitted to intersect except at the endpoints. A closed closed curve can be included in this framework by picking two points on the curve and break the curve into two simple curves sharing two distinct endpoints.} \)

Let \( \Delta_0, C_s > 0, \gamma_1 \in (0, 1) \) and \( 0 < \Delta < \ell_s/2 \), where \( \ell_s = \ell_s(\Delta) \sim \Delta^{\gamma_s} \). The constants \( C^k, \ell^k \) have the same restrictions as \( C_s, \ell_s \), respectively. Let \( \Gamma^k, \Delta = \Gamma^k \setminus \bigcup_s B_{\ell_s}(z_s), \Gamma^k, \Delta, s = (B_{\ell_s}(z_s) \cap \Gamma^k) \setminus B_{\ell_s/2}(z_s). \)

We require the following properties to hold for \( \Delta \leq \Delta_0 \):

1. If \( z_s \) is not an endpoint of \( \Gamma^k \), then \( \text{dist}(z_s, \Gamma^k) \geq \ell_s \).
2. If \( z_s \) is an endpoint of \( \Gamma^k \), then \( \Gamma^k \) only intersects \( \partial B_{\ell_s}(z_s) \) in a single point.
3. \( \text{dist}(\Gamma^k, \Delta, \Gamma^k, \Delta) \geq (1 - \delta_{k,k'})\Delta. \)
4. Suppose that \( \Gamma^k \) has endpoints \( z_{s_1} \) and \( z_{s_2} \). There is a partition of \( \Gamma^k \) into disjoint consecutive arcs \( A^k, \Delta^1, \ldots, A^k, \Delta^r \subset \Gamma^k, \Delta \) satisfying the following properties:
   (a) \( \frac{1}{C^k} \ell^k \leq \text{diam}(A^k, \Delta) \leq C^k \ell^k \).
   (b) \( \text{dist}(A^k, \Delta, A^k, \Delta') \geq (1 - \delta_{i,i'})\Delta. \)
   (c) Each \( A^k, \Delta_i \) can be broken up into consecutive subarcs \( A^k, \Delta_{ij} \) each satisfying \( \text{diam}(A^k, \Delta_{ij}) \leq C^k \Delta \) and \( \text{dist}(A^k, \Delta_{ij}, A^k, \Delta_{ij'}) \geq (1 - \delta_{j,j'})\Delta. \)
   (d) If \( z_s \) is an endpoint of \( \Gamma^k \) then there is a unique value \( i_s^k \) of \( i \) and \( j_s^k \) of \( j \) so that \( A^k, \Delta_{ij} \) is within a distance of \( \Delta \) from \( B_{\ell_s}(z_s) \). One of the endpoints of this subarc will then be on \( \partial B_{\ell_s}(z_s) \).

How small that \( \Delta \) needs to be will depend on how close the \( \Gamma^k \) are near and away from \( z_j \). It is necessarily the case that if there are many \( \Gamma^k \) with endpoint \( z_s \), then \( l/\Delta \) must be large. This is illustrated by Figure 4 below, in particular how visually near the intersection point one can compare the size of the black “dot” formed by the intersection of line segments depicted with positive width.

![Figure 5: For a given \( r > 0 \), compare how far one has to move from the point of intersection so that each line is at least a distance of \( r \) from the other lines.](image)

We then have the following:
Figure 6: \( \Gamma^{k,\Delta} \) is depicted in blue, \( \partial B_{l_i}(z_j) \) in red, and \( \Gamma^{k,\Delta} \cap B_{l_i}(z_s) \) in black. Although not all end points are shared by multiple curves \( \Gamma^k \) in general, it is this case in the illustration above.

**Proposition 18.3.** Let \( N \) be normal with spectrum in \( \Gamma \) satisfying \((\ast)\) and \( A \) self-adjoint. Then if \( A,N \) are almost commuting, they are nearly commuting.

**Proof.** Let \( \delta = \|[A,N]\| \). We apply the argument earlier to obtain \( H \) finite range with respect to \( N \) of distance \( \Delta \sim \delta^{1/2} \) such that \( \|A - H\| \leq \text{Const.}\delta^{1-\gamma} \), \( \|[H,N]\| \leq \text{Const.}\delta \). Recall \((\ast)\).

Let \( B^{k,\Delta}_i, V^{k,\Delta}_{i,j} \) be the range of the spectral projection of \( N \) onto \( A^{k,\Delta}_i, A^{k,\Delta}_i \), respectively. We then obtain \( V^{k,\Delta}_{i,j} \) which are subspaces of the \( B^{k,\Delta}_i \), satisfy \( V^{k,\Delta}_{i,j} \subset W^{k,\Delta}_{s} \perp V^{k,\Delta}_{i,j} \), and are almost invariant under \( P_{B^{k,\Delta}_i} H P_{B^{k,\Delta}_i}^\ast \). Let \( W^{k,\Delta}_{i,j} = B^{k,\Delta}_i \ominus V^{k,\Delta}_{i,j} \).

We now need to deal with the subspace \( B^{\Delta,s} \) projected onto by \( E_{B_{l_i}}(z_s)(N) \). The issue is that because \( z_s \) can be the endpoint of multiple curves \( \Gamma^k \), we have to use a slightly different approach to form almost invariant subspaces of \( B^{\Delta,s} \) so that everything works out. This is the only part of the proof where one uses a “two dimensional” grouping of eigenvalues of \( N \) (using “polar coordinates”).

Fix an endpoint \( z_s \). Let \( S_1 = B_{\ell_s/2}(z_s), S_k^0 = \Gamma^k \cap B_{(i-1)\Delta + \ell_s/2}(z_s) \setminus B_{(i-2)\Delta + \ell_s/2}(z_s), S_{r,s}^1 = \Gamma^k \cap B_{s}(z_s) \setminus B_{(r-1)\Delta + \ell_s/2}(z_s), \) where \( \ell_s/2 - 1 < r \Delta,s \Delta \leq \ell_s/2 \). What we do is break \( B^{\Delta,s} \) into orthogonal subspaces \( V^{1,\Delta,s}, V^{2,\Delta,s}, \ldots, V^{r,\Delta,s} \) by letting \( E_{S_1}(N) \) project onto \( V^{1,\Delta,s} \) and \( E_{S_{r,s}}(N) \) project onto \( V^{r,\Delta,s} \) for \( 1 \leq i \leq r \). Consider \( H^{\Delta,s} = P_{B^{\Delta,s}} H P_{B^{\Delta,s}}^\ast \) as a block matrix with respect to the subspaces \( V^{1,\Delta,s}, V^{r,\Delta,s}, \ldots, V^{r,\Delta,s} \).

For each \( k \), we restrict \( H \) to \( B^{k,\Delta,s} = V^{1,\Delta,s} \oplus \bigoplus_{i=2}^{r} V^{i,\Delta,s} \) to obtain \( H^{k,\Delta,s} \). \( H^{k,\Delta,s} \) is tridiagonal with respect to the subspaces \( V^{1,\Delta,s}, V^{r,\Delta,s}, \ldots, V^{r,\Delta,s} \) which correspond to a set covering \( \Gamma^k \cup S_1 \) in \( B_{\ell_s}(z_s) \). Then there is a subspace \( W^{\Delta,s}_k \subset B^{k,\Delta,s} \) that is almost invariant under \( H^{k,\Delta,s} \) such that \( V^{k,\Delta,s}_{l,s} \subset W^{\Delta,s}_k \perp V^{1,\Delta,s}_1 \).

Now, \( P_{W^{\Delta,s}_k} \leq E_{\Gamma^{k,\Delta,s}}(N) \) so not only are the subspaces \( W^{\Delta,s}_k \) orthogonal for different values of \( k \), but \( H^{\Delta,s} \) maps \( W^{\Delta,s}_k \) into \( W^{\Delta,s}_k \uplus V^{1,\Delta,s} \) due to the assumption on the distance between the \( \Gamma^{k,\Delta,s} \) for different values of \( k \). Hence, \( H^{\Delta,s} \) is equal to \( H^{k,\Delta,s} \) on \( W^{\Delta,s}_k \) and \( H^{\Delta,s}(W^{\Delta,s}_k) \) is perpendicular to \( W^{\Delta,s}_{k'} \) for \( k' \neq k \) since \( W^{\Delta,s}_{k'} \perp V^{1,\Delta,s}_1 \) by
The point $z_s$ is an endpoint for $\Gamma^1, \Gamma^2, \Gamma^3$, and $\Gamma^4$. The small tick marks are the intersections of circles centered at $z_s$ of radii $i\Delta$. The almost invariant subspaces $W_{\Delta,s}^{\pm}$ are illustrated in red and $W_{\Delta,s}^{\perp}$ in blue. Although these subspaces do not necessarily lie entirely within the span of the illustrated eigenspaces of $N$, $W_{\Delta,s}^{\perp}$ contains all eigenspaces of $N$ within distance $l_s/2$ of $z_s$ and $W_{\Delta,s}^{\pm}$ contains the eigenspaces depicted by part of the (red) spectrum near the large tick marks.

construction. Then we find that $W_{\Delta,s}^{\pm}$ are almost invariant under $H_{\Delta,s}$ because

$$(1 - P_{W_{\Delta,s}^{\pm}})H_{\Delta,s}P_{W_{\Delta,s}^{\pm}} = P_{\mathcal{B}_{\Delta,s}^{\pm} \oplus W_{\Delta,s}^{\pm}}H_{\Delta,s}P_{W_{\Delta,s}^{\pm}}$$

has small norm.

Let $W_{\Delta,s}^{\perp} = \mathcal{B}_{\Delta,s}^{\perp} \oplus \bigoplus W_{\Delta,s}^{\pm}$. Because $H_{\Delta,s}$ is Hermitian, we have broken $\mathcal{B}_{\Delta,s}^{\pm}$ into almost invariant subspaces $W_{\Delta,s}^{\pm}$ into almost invariant subspaces $W_{\Delta,s}^{\pm}$ and $W_{\Delta,s}^{\perp}$.

We now form the new basis of subspaces $\tilde{\mathcal{B}}$. This is a simple process, but its explanation is complicated by the fact that because of parity issues, we cannot just say “join $W$’s with $W^{\perp}$’s” as before. We let the subspaces $\tilde{\mathcal{B}}$ include three types of subspaces. The first subspaces that we include are $W_{\Delta,s}^{\perp}$. The other “$W$” subspaces contain a “$V$” subspace which has to be matched with its neighboring “$V$” subspace. One of the subspaces $W_{i}^{k,\Delta}, W_{i}^{k,\Delta,\perp}$ contains that subspace $W_{i}^{k,\Delta}, W_{i}^{k,\Delta,\perp}$, call this subspace $W_{i}^{k,\Delta,\perp}$. We then include the subspace $W_{i}^{k,\Delta,\perp} \oplus W_{k}^{\Delta,s}$. We also include the direct sums of the remaining consecutive subspaces $W_{i}^{k,\Delta,\perp} \oplus W_{i+1}^{\Delta,s}$.

Having defined the subspaces $\tilde{\mathcal{B}}$, we proceed to defining the nearby commuting matrices. We define $N'$ to be the normal matrix that is block identity picking an eigenvalue from the arcs (or union of arcs) that we used to make the spaces $\tilde{\mathcal{B}}$. We then have $\|N - N'\| \leq 4\max_s C_s \ell_s$ and $N'$ commutes with $H' = \sum_{\mathcal{B}} P_{\mathcal{B}} H P_{\mathcal{B}}$ which is close to $H$. \qed

19 Macroscopic Observables

The notion of almost commuting operators associated to observables being near actually commuting operators is discussed and used in a paper by von Neumann, translation provided in [49]. A specific passage in the beginning of the article states:
Still, it is obviously factually correct that in macroscopic measurements the coordinates and momenta are measured simultaneously — indeed, the idea is that that becomes possible through the inaccuracy of the macroscopic measurement, which is so great that we need not fear a conflict with the uncertainty relations.

We believe that the following interpretation is the correct one: in a macroscopic measurement of coordinate and momentum (or two other quantities that cannot be measured simultaneously according to quantum mechanics), really two physical quantities are measured simultaneously and exactly, which however are not exactly coordinate and momentum. They are, for example, the orientations of two pointers or the locations of two spots on photographic plates—and nothing keeps us from measuring these simultaneously and with arbitrary accuracy, only their relation to the really interesting physical quantities \(q_k\) and \(p_k\) is somewhat loose, namely the uncertainty of this coupling required by the laws of nature corresponds to the uncertainty relation[.]

Von Neumann then proceeds to discuss various aspects of this. Directly related to this topic of quantum statistical mechanics is the idea of macroscopic observables (“macroscopic measurements” in the above quote) as defined and discussed in Section II B. of [50].

Two almost commuting self-adjoint matrices are nearby actually commuting matrices. However, such a definitive statement is not true for more than two almost commuting matrices under the operator norm as discussed in previous sections. It is true under other norms as discussed in Section 2. Moreover, Ogata in [7] provided a proof for the special case of arbitrarily many macroscopic observables with the operator norm “because of their nice thermodynamic structure”.

We phrase the result in terms of the operators \(T_N : M_n(\mathbb{C}) \rightarrow M_{n^N}(\mathbb{C})\) defined by:

\[
T_N(A) = \frac{1}{N} \sum_{k=0}^{N-1} I_n^{(N-1-k)} \otimes A \otimes I_n^{(N-1-k)}.\]

Because \(I_n^{(N-1-j)} \otimes A \otimes I_n^{(N-1-k)} \otimes B \otimes I_n^{(N-1-k)}\) equals 0 if \(j \neq k\) and equals \(I_n^{(N-1-k)} \otimes [A, B] \otimes I_n^{(N-1-k)}\) if \(j = n\), we see that

\[
[T_N(A), T_N(B)] = \frac{1}{N} T_N([A, B]). \tag{34}
\]

Also, for \(U \in M_n(\mathbb{C})\) unitary,

\[
T_N(UAU^*) = U^{\otimes N} T_N(A)(U^*)^{\otimes N}. \tag{35}
\]

With that said, we have that for any collection of matrices, applying \(T_N\) to them as \(N \rightarrow \infty\) gives a sequence of collections of almost commuting matrices. Ogata’s result implies:

**Theorem 19.1.** For \(A_1, \ldots, A_m \in M_n(\mathbb{C})\) self-adjoint, let \(H_{i,N} = T_N(A_i)\). Then there are commuting matrices \(Y_{i,N} \in M_{n^N}(\mathbb{C})\) so that \(\|H_{i,N} - Y_{i,N}\| \rightarrow 0\) as \(N \rightarrow \infty\).
Remark 19.2. This result is nonconstructive. Note that Ogata’s result in [7] is only stated in terms of odd $N$, in slightly different notation. However, because

$$T_{N+1}(A) = \frac{N}{N+1} (T_N(A) \otimes I_n + I_n \otimes T_N(A)),$$  

we can set

$$Y_{i,2N} = \frac{2N-1}{2N} (Y_{i,2N-1} \otimes I_n + I_n \otimes Y_{i,2N-1})$$

so that since there are nearby commuting matrices for $N$ odd, we have the same result for $N$ even.

We now list some properties of $T_N$. When $A$ is diagonalizable, we see that

$$\sigma(T_N(A)) = \frac{1}{N} \sum_{k=0}^{N-1} \sigma(A).$$  

Thus, the spectrum of $T_N(A)$ is a discrete approximation of the convex hull of $\sigma(A)$. This shows that the methods discussed in Section 18 will not apply in general for $m \geq 3$. As a consequence of Equation (37), $||T_N(A)|| = ||A||$ when $A$ is normal and in general $||T_N(A)|| \leq ||A||$ by definition. Applying

$$||A|| \leq ||\text{Re} A|| + ||\text{Im} A|| \leq 2||A||$$

to $T_N(A)$, we see that

$$\frac{1}{2}||A|| \leq ||T_N(A)|| \leq ||A||.$$

Consequently, $T_N$ is injective.

Two questions we ask are:

1. What properties of $T_N$ imply that there are nearby commuting matrices?
2. How much special structure for the $Y_{i,N}$ can be ensured?

Discussing some of the properties of $T_N$, note that

$$||T_N(A) \otimes I_n - I_n \otimes T_N(A)|| = \frac{1}{N} ||A \otimes I^\otimes N - I^\otimes N \otimes A|| = O(1/N).$$

Suppose that we define $F_N(A) \in M_{k_N}(\mathbb{C})$ to satisfy

$$||F_N(A) \otimes I_{k_N} - I_{k_N} \otimes F_N(A)|| = o(1)$$

as $N \to \infty$. If the $F_N(A_i)$ are self-adjoint, they are then unitarily equivalent to $D_i = \text{diag}(a_{N,i}^1, \ldots, a_{N,k_N}^i)$. Because

$$\max_{i,j} |a_{N,i}^j - a_{N,j}^i| = ||D \otimes I_{k_N} - I_{k_N} \otimes D|| = o(1),$$

defining $Y_N^i = a_{N,1}^i I_{k_N}$ shows that the $F_N(A_i)$ are nearly commuting.
Although this implies that they are almost commuting, there is a more direct proof of this:

\[ \| [F_N(A), F_N(B)] \| = \| [F_N(A), F_N(B)] \otimes I_{kN} \| = \| [F_N(A) \otimes I_{kN}, F_N(B) \otimes I_{kN}] \| \\
= \| [I_{kN} \otimes F_N(A), F_N(B) \otimes I_{kN}] \| + o(1) = o(1). \]

This, of course, was a trivial example. The key property from (40) being used was that the identity matrix \( I_{kN} \) has the same dimension as \( F_N(A) \). If you write out an eigenbasis \( v_1, \ldots, v_2 \) for \( F_N(A) \) then the basis \( v_i \otimes v_j \) (with the lexicographic ordering) causes the eigenvalues for \( F_N(A) \otimes I_{kN} \) to be \( a_1, \ldots, a_2, \ldots, a_2, \ldots, a_{kN}, \ldots, a_{kN} \) and the eigenvalues of \( I_{kN} \otimes F_N(A) \) to be \( a_1, a_2, \ldots, a_{kN}, a_1, \ldots, a_{kN}, \ldots, a_{kN} \). Each sublist has length \( k_N \). So, we are listing each of the \( k_N \) eigenvalues \( k_N \) times, but in different ways.

However, if the number \( n \) of repetitions is much smaller than the number \( k_N \) of eigenvalues then the condition

\[ \| F_N(A) \otimes I_n - I_n \otimes F_N(A) \| = o(1) \]

is not as strong. In particular, the condition that we have for macroscopic observables is

\[ \| T_N(A) \otimes I_n - I_n \otimes T_N(A) \| = O(1/N). \]

Now, in what we will see is relevant later, we can write this as

\[ \| T_N(A) \otimes I_n - I_n \otimes T_N(A) \| = O(1/\log(k_N)) \]  \hspace{1cm} (41)

because \( k_N = n^N \).

Although we already know that the \( T_N(A_i) \) are almost commuting, if we require slightly more than (41), then we can obtain almost commutativity from this equality. Consider

\[ \| F_N(A) \otimes I_n - I_n \otimes F_N(A) \| = o(1/ \log(k_N)). \]

We only look at \( k_N = n^{cN} \). With this condition, we note that

\[ F_N(A) \otimes I_{kN} = F_N(A) \otimes I_n^{\otimes cN} = I_n^{\otimes cN} \otimes F_N(A) + c_N o(1/ \log(k_N)) = I_{kN} \otimes F_N(A) + o(1). \]

So as before, \( F_N(A_1), \ldots, F_N(A_m) \) almost commute.

An open question is if an inequality like (42) or even (41) along with perhaps the other properties listed above that \( T_N \) satisfies is enough to ensure that that \( F_N(A_1), \ldots, F_N(A_m) \) nearly commute.

We now attempt to address the second question that we posed above. We first note that if \( v_1, \ldots, v_k \) is a basis for \( M_n(\mathbb{C}) \) then \( v_{i_1} \otimes \cdots \otimes v_{i_N} \) form a basis for \( M_n(\mathbb{C})^{\otimes N} \). There is a natural action of an element \( \sigma \) of the symmetric group \( S^N \) on the basis by:

\[ \sigma \cdot (v_{i_1} \otimes \cdots \otimes v_{i_N}) = v_{\sigma(i_1)} \otimes \cdots \otimes v_{\sigma(i_N)}. \]

We then obtain a representation of \( S^N \). For any matrix \( A \in M_n(\mathbb{C}) \), \( T_N(A) \) commutes with the representation.

A natural question that we might ask is if we can ensure that the \( Y_{1N} \) also commute with the representation. If we are only looking at \( m = 2 \) (so \( i = 1, 2 \)) then we can apply
Lin’s theorem to obtain this result as follows. Because \( T_N(A_i) \) commute with the representation we see that all elements of the unital \( C^* \)-algebra \( \mathcal{C}_N = C^*(\{T_N(A_i)\}_{i=1}^n, I_n) \) commute with the representation. By the characterization of finite dimensional \( C^* \)-algebras (see Theorem III.1.1 of [53]), we see that \( \mathcal{C}_N \cong \bigoplus_i M_{k_i}(C) \) isometrically.

Then there are self-adjoint matrices \( A_{i,N,\ell} \) such that \( T_N(A_i) \sim \bigoplus_\ell A_{i,N,\ell} \).

We know that \( A_{1,N,\ell}, A_{2,N,\ell} \) almost commute with commutator bounded in norm by \( \|T_N(A_1), T_N(A_2)\| \), so by Lin’s theorem we obtain nearby commuting matrices \( A'_{i,N,\ell} \). Then we obtain commuting \( Y_{i,N} \cong \bigoplus_\ell A'_{i,N,\ell} \) that are close to \( T_N(A_i) \), commute, and lie in \( \mathcal{C}_N \). Hence, we showed that there are nearby commuting matrices \( Y_{i,N} \) that commute with the representation.

Consider von Neumann’s comments which occur just before the passage that is quoted above about phase space for the purposes of Quantum statistical mechanics.

But the phase space cannot be formed in quantum mechanics, since a coordinate \( q_k \) and the corresponding momentum \( p_k \) are never simultaneously measurable; instead, their probable errors (spreads) \( \Delta q_k \) and \( \Delta p_k \) are always related according to the uncertainty relation \( \Delta q_k \Delta p_k \geq \hbar/2 \). Moreover, it is impossible to specify, for a state of the system, two intervals \( I, J \) so that, with certainty, \( q_k \) lies in \( I \) and \( p_k \) in \( J \) (even if the product of their lengths is much bigger than \( \hbar/2 \)) – thus, not only the continuous phase space but also a discrete partition thereof into cells is meaningless!

In accordance with the above quote, we do not provide nearby commuting matrices \( Y_{1,N}, Y_{2,N} \) that are able to respect the eigenspaces of both \( T_N(A_1) \) and \( T_N(A_2) \). However, we can assure that one of the \( Y_{i,N} \) can respect the eigenspaces of \( T_N(A_i) \).

At the end of Section 1.1., von Neumann discusses macroscopic measurements and a partitioning of the spectrum which we formulate in terms of our problem as follows. We can partition the spectrum of the operator \( T_N(A_1) \) into disjoint intervals which we can view as the different results of macroscopic measurements because only so much accuracy can be attained for such a system. What we look for is that the eigenvectors of \( Y_{1,N} \) of eigenvalue \( \lambda \) are expressed as the span of only eigenvectors of \( T_N(A_1) \) with eigenvalues near \( \lambda \). This means that a measurement of our nearby observable corresponds to a superposition of states with nearby values for the original operator \( T_N(A_1) \). Naturally, by the uncertainty principle, we do not expect the same to be true for \( Y_{2,N} \) for the same choice of nearby commuting matrices.

We focus in on the above construction of nearby commuting operators \( Y_{i,N} \) that commute with the representation of the symmetric group. The spectrum of \( T_N(A_i) \) can be expressed as \( \bigcup_\ell \sigma(A_{i,N,\ell}) \) by the \( C^* \)-algebra isomorphism. Then instead of “just” applying Lin’s theorem, we can apply a proof of Lin’s theorem using \( (Q') \) as described in Section 7.3 (see Remark 7.3) with \( A = A_{1,N,\ell} \) and \( B = A_{2,N,\ell} \) to obtain \( A'_{i,N,\ell} \) where the eigenspaces of \( A'_{i,N,\ell} \) are subspaces of local spans of eigenspaces of \( A_{1,N,\ell} \). Then form \( Y_{i,N} \cong \bigoplus_\ell A'_{i,N,\ell} \). If we partition the eigenvalues of \( Y_{1,N} \) into disjoint intervals, then this corresponds to nonconsecutively disjoint intervals covering the spectrum of \( T_N(A_1) \) with the property we discussed above.

An open question is if any of the above structure for the \( Y_{i,N} \) can be maintained when \( m \geq 3 \).

One way to reformulate the problem is to consider \( \mathcal{A}_N = T_N(M_n(C)) \subset M_{n^n}(C) \). To show that given any finite collection of Hermitian matrices \( A_1, \ldots, A_m' \), the \( T_N(A_i) \)
have nearby commuting counterparts, it suffices to choose a basis $A_1, \ldots, A_m$ for $M_n(\mathbb{C})$ of Hermitian contractions. If we have a contraction $\tilde{A} \in \mathcal{A}_N$, then there is a unique $A \in M_n(\mathbb{C})$ so that $T_N(A) = \tilde{A}$ and by Equation (38), $\|A\| \leq 2$. Because the $A_i$ are a basis, we have $A = \sum_i c_i A_i$ and by the finite dimensionality, $\max_i |c_i| \leq C = C(n)$. Thus setting $Y = \sum_i c_i Y_{i,N}$, we obtain that $Y$ lies in the span of the $Y_{i,N}$ and $\|Y - T_N(A)\| \leq nC \max_{1 \leq i \leq N} \|T_N(A_i) - Y_{i,N}\|$, which converges to zero as $N \to \infty$.

Ogata’s result assures that as $N \to \infty$, $\mathcal{A}_N$ grows closer to some commuting subspace of $M_n(\mathbb{C})$. Expanding what was said in Section 17, finding the matrices $Y_{i,N}$ is equivalent to finding a minimizer (or an approximate minimizer) to

$$\min_{U \in U(n^N)} \max_{A \in B_{\mathcal{A}_N}} \|UAU^* - \text{diag}(UAU^*)\| \tag{43}$$

which is equivalent to

$$\min_{U \in U(n^N)} \max_{1 \leq i \leq m} \|UT_N(A_i)U^* - \text{diag}(UT_N(A_i)U^*)\|. \tag{44}$$

Continuing the previous approach, because $T_N$ is an injection, we could consider the normed Lie algebra $\mathcal{A}_N = (M_n(\mathbb{C}), [-, -]_N, \|\|_N)$, where $A \in \mathcal{A}_N$ is identified with $T_N(A)$. In that way, by (38), $\|A\|_N$ is equivalent to $\|A\|$ and $[-, -]_N = \frac{1}{N}[-, -]$. So, we can easily see that because the Lie bracket converges to zero as $N \to \infty$, all bounded sets of elements are almost commuting. However, because for each $N$, $\| - \|_N$ is equivalent to $\| - \|$, there are not any matrices that do not commute that will be nearby almost commuting matrices in $\mathcal{A}_N$. This shows that to find any nearby commuting matrices, we would need to embed the Lie algebra into a larger space. An open question is if this larger space can be $\mathcal{C}_N$ as a strictly smaller $C^*$-subalgebra of $M_{nN}(\mathbb{C})$.

ACKNOWLEDGEMENTS. The author would like to thank Eric A. Carlen for introducing the problem to the author and continued guidance during the writing and revising of this paper and Matthew Hastings for some corrections of and clarifications concerning [1].

This research was partially supported by NSF grant DMS 1501007.
References

[1] M. Hastings. Making Almost Commuting Matrices Commute. Version 4. https://arxiv.org/abs/0808.2474

[2] M. Hastings. Making Almost Commuting Matrices Commute. Commun. Math. Phys. 291, 321–345 (2009)

[3] Matthew Hastings. Personal Communication.

[4] H. Lin. Almost commuting self-adjoint matrices and applications. Fields. Inst. Commun. 13, 193 (1995).

[5] K. Davidson. Almost Commuting Hermitian Matrices. Math. Scand. 56 (1985), 222-240.

[6] C. Pearcy and A. Shields. Almost Commuting Matrices. Journal of Functional Analysis. 33, 332-338 (1979).

[7] Y. Ogata. Approximating macroscopic observables in quantum spin systems with commuting matrices. Journal of Functional Analysis. Vol. 264, Issue 9, 1 May 2013, pp 2005-2033.

[8] T. Loring and A. Sørensen. Almost-Commuting Self-Adjoint Matrices - The Real and Self-Dual Cases. Reviews in Mathematical Physics, 28(07): 1650017, 2016.

[9] M. Hastings and T. Loring. Almost commuting matrices, localized Wannier functions, and the quantum Hall effect. J. Math. Phys. 51, 015214 (2010). https://doi.org/10.1063/1.3274817

[10] I. Kachkovskiy and Y. Safarov. Distance to Normal Elements in C*-Algebras of Real Rank Zero. https://arxiv.org/abs/1403.2021

[11] T. Osborne. Almost Commuting Unitaries with Spectral Gap are Near Nearly Commuting Matrices. https://arxiv.org/abs/0809.0602

[12] T. Loring. K-theory and Asymptotically Commuting Matrices. Can. J. Math., Vol. XL, No. 1, 1988, pp. 197-216.

[13] R. Exel and T. Loring. Almost Commuting Unitary Matrices. Proc. Amer. Math. Soc. Vol. 106, No. 4, 1989.

[14] R. Exel and T. Loring. Invariants of Almost Commuting Unitaries.

[15] H. Lin. Almost Commuting Unitaries and Classification of Purely Infinite Simple C*-Algebras. Journal of Functional Analysis, 10 May 1998, Vol.155(1) pp.1-24.

[16] P. Friss and M. Rørdam. Almost commuting self-adjoint matrices - a short proof of Huaxin Lin's theorem. J. reine angew. Math 479 (1996). 121-131.

[17] S. Szarek. On Almost Commuting Hermitian Operators. Rocky Mountain Journal of Mathematics. Vol. 20, No. 2, Spring 1990.

[18] C. Davis. The Rotation of Eigenvectors by a Perturbation. Journal of Mathematical Analysis and Applications 6, 159-173 (1963) https://doi.org/10.1016/0022-247X(63)90001-5

[19] M. Said. Almost Commuting Elements in Non-Commutative Symmetric Operator Spaces. Dissertation.
REFERENCES

[20] C. Jordan. *Essai sur la géomécatrie à n dimensions.*

[21] J. Dixmier. *Position relative de deux variétés linéaires fermées dans un espace de Hilbert.*

[22] P. Halmos. *Two Subspaces.*

[23] P. Halmos. *Some unsolved problems of unknown depth about operators on Hilbert space* Proceedings of the Royal Society of Edinburgh, 76A, 67-76, 1976. [https://doi.org/10.1017/S0308210500019491](https://doi.org/10.1017/S0308210500019491)

[24] P. Rosenthal. *Are Almost Commuting Matrices Near Commuting Matrices?* The American Mathematical Monthly, Vol. 76, No. 8 (Oct., 1969), pp. 925-926.

[25] W. Luxemburg and R. Taylor. *Almost Commuting Matrices are near Commuting Matrices.*

[26] J. Bastian and K Harison. *Subnormal Weighted Shifts and Asymptotic Properties of Normal Operators* Proc. Amer. Math. Soc. 42 (1974), 475-479. [https://doi.org/10.1090/S0002-9939-1974-0380491-X](https://doi.org/10.1090/S0002-9939-1974-0380491-X)

[27] I. Berg and C. Olsen. *A note on Almost Commuting Matrices.* Michigan Math. J. Volume 21, Issue 4 (1975), 377-383.

[28] D. Voiculescu. *Asymptotically commuting finite rank unitary operators without commuting approximants.* Acta Sci. Math. (Szeged) 45:1-4(1983), 429-431 1983.

[29] K. Davison and S. Szarek. Chapter 8 Local Operatorn Theory, Random Matrices, and Banach Spaces of *Handbook of the Geometry of Banach Spaces Vol 1.* 2001.

[30] D. Enders and T. Shulman. *Almost Commuting Matrices, Cohomology, and Dimension.*

[31] I. Berg. *On Approximation of Normal Operators by Weighted Shifts.*

[32] S. Demko, W. Moss, and P. Smith. *Decay Rates for Inverses of Band Matrices.* Mathematics of Computation, 43(168), 491-499. (1984) doi:10.2307/2008290

[33] O. Bratteli and D. Robinson. *Operator Algebras and Quantum Statistical Mechanics.* Volume 1. 1979.

[34] A. B. Aleksandrov and V. Peller. *Operator Lipschitz functions.* 2016 Russ. Math. Surv.71 605.

[35] A.B. Aleksandrova, V.V. Peller. *Estimates of operator moduli of continuity.* Journal of Functional Analysis 261 (2011) 2741–2796.

[36] A. B. Aleksandrov and V. V. Peller. *Operator Hölder-Zygmund Functions.* [https://arxiv.org/abs/0907.3049](https://arxiv.org/abs/0907.3049)

[37] C. Davis and W. Kahan. *The Rotation of Eigenvectors by a Perturbation. III.* SIAM J. Numer. Anal., 7(1), 1–46. [https://doi.org/10.1137/0707001](https://doi.org/10.1137/0707001)

[38] R. Bhatia, C. Davis, and A. McIntosh. *Perturbation of Spectral Subspaces and Solution of Linear Operator Equations.* Linear Algebra and its Applications. 52/53:45-67 (1983) [https://doi.org/10.1016/0024-3795(83)80007-X](https://doi.org/10.1016/0024-3795(83)80007-X)

[39] R. Bhatia and P. Rosenthal. *How and Why to Solve the Operator Equation AX − XB = Y.* [https://doi.org/10.1112/S0024609396001828](https://doi.org/10.1112/S0024609396001828)
REFERENCES

[40] R. Bhatia, C. Davis and P. Koosis. An extremal problem in Fourier analysis with applications to operator theory. https://doi.org/10.1016/0022-1236(89)90095-5

[41] L. Glebsky. Almost commuting matrices with respect to normalized Hilbert-Schmidt norm. https://arxiv.org/abs/1002.3082

[42] C. Spengler, M. Huber, and B. C. Hiesmayr. A composite parameterization of unitary groups, density matrices and subspaces. https://arxiv.org/abs/1004.5252

[43] Spengler, M. Huber, and B. C. Hiesmayr. Composite parameterization and Haar measure for all unitary and special unitary groups.

[44] C. Spengler. Composite parameterization of unitary groups. https://www.mathworks.com/matlabcentral/fileexchange/30990-composite-parameterization-of-unitary-groups

[45] M. Benzi and G. Golub. Bounds for the Entries of Matrix Functions with Applications to Preconditioning. BIT Numerical Mathematics (1999) 39: 417. https://doi.org/10.1023/A:1022362401426

[46] S. Pironio. et al. Device-independent quantum key distribution secure against collective attacks. 2009 New J. Phys.11 045021

[47] D. Nagaj, P. Wocjan, Y. Zhang. Fast Amplification of QMA. https://arxiv.org/abs/0904.1549

[48] A. Prakash. Quantum Algorithms for Linear Algebra and Machine Learning. Dissertation. https://www2.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS-2014-211.html

[49] J. von Neumann. Proof of the ergodic theorem and the H-theorem in quantum mechanics Translation of: Beweis des Ergodensatzes und des H-Theorems in der neuen Mechanik Eur. Phys. J. H35, 201–237 (2010).

[50] D. Poulin. Macroscopic observables. Physical Review A 71, 022102 (2005).

[51] M. Andelić and C. da Fonseca. Sufficient conditions for positive definiteness of tridiagonal matrices revisited. https://doi.org/10.1007/s11117-010-0047-y

[52] C. Davis. Generators of the Ring of Bounded Operators. Proc. Amer. Math. Soc. Vol. 6, No. 6 (Dec., 1955), pp.970-972.

[53] K. Davidson. C*-Algebras by Example. AMS. 1996.

[54] B. Blackadar. Operator Algebras - Theory of C*-Algebras and von Neumann Algebras. Springer-Verlag. Berlin Heidelberg. 2006.

[55] L. Grafakos. Classical Fourier Analysis. 2nd. Edition. 2008.

[56] E. Stein and R. Shakarchi. Complex Analysis. Princeton University Press. 2003.