Sparse Pseudospectral Shattering

Rikhav Shah* UC Berkeley
Nikhil Srivastava† UC Berkeley
Edward Zeng NYU

November 27, 2024

Abstract

The eigenvalues and eigenvectors of nonnormal matrices can be unstable under perturbations of their entries. This renders an obstacle to the analysis of numerical algorithms for non-Hermitian eigenvalue problems. A recent technique to handle this issue is pseudospectral shattering [BGVKS23], showing that adding a random perturbation to any matrix has a regularizing effect on the stability of the eigenvalues and eigenvectors. Prior work has analyzed the regularizing effect of dense Gaussian perturbations, where independent noise is added to every entry of a given matrix [BGVKS24, BGVKS23, BKMS21, JSS21].

We show that the same effect can be achieved by adding a sparse random perturbation. In particular, we show that given any $n \times n$ matrix M of polynomially bounded norm: (a) perturbing $O(n \log^2(n))$ random entries of M by adding i.i.d. complex Gaussians yields $\log \kappa_V(A) = O(\mathrm{poly}\log(n))$ and $\log(1/\eta(A)) = O(\mathrm{poly}\log(n))$ with high probability; (b) perturbing $O(n^{1+\alpha})$ random entries of M for any constant $\alpha > 0$ yields $\log \kappa_V(A) = O_\alpha(\log(n))$ and $\log(1/\eta(A)) = O_\alpha(\log(n))$ with high probability. Here, $\kappa_V(A)$ denotes the condition number of the eigenvectors of the perturbed matrix A and $\eta(A)$ denotes its minimum eigenvalue gap.

A key mechanism of the proof is to reduce the study of $\kappa_V(A)$ to control of the pseudospectral area and minimum eigenvalue gap of A, which are further reduced to estimates on the least two singular values of shifts of A. We obtain the required least singular value estimates via a streamlining of an argument of Tao and Vu [TV07] specialized to the case of sparse complex Gaussian perturbations.

1 Introduction

A central question in numerical analysis is "how do the eigenvalues and eigenvectors of a matrix behave under perturbations of its entries?" For normal matrices, the eigenvalues are $1-$Lipschitz functions of the entries, and the eigenvectors are stable under perturbations if the minimum eigenvalue gap is large. This fact is essential to the rapid convergence and rigorous analysis of algorithms for the Hermitian eigenvalue problem and its cousins.

For nonnormal matrices, two related difficulties appear: *non-orthogonality* of the eigenvectors and *spectral instability*, i.e. high sensitivity of the eigenvalues to perturbations of the matrix entries. Non-orthogonality slows down the convergence of iterative algorithms (such as the power method) and spectral instability makes it difficult to rigorously reason about convergence in the presence of roundoff error. The main tool used to surmount these difficulties in recent years is smoothed analysis, i.e., adding a small random perturbation to the input and solving the perturbed problem, incurring a small backward error[1]. Specifically it was shown in [BGVKS23] that adding small i.i.d. complex Gaussian random variables to each entry of a matrix produces a matrix with well-conditioned eigenvectors and a large eigenvalue gap, a phenomenon termed "pseudospectral shattering." This was then generalized to other random variables in [BGVKS24, JSS21, EJ24], and is

*Supported by NSF CCF-2420130
†Supported by NSF CCF-2420130

[1] i.e., the algorithm produces an exact solution to a nearby problem, rather than producing an approximate solution to the given problem, which is called forward error.

currently an essential mechanism in all of the known convergence results about diagonalizing arbitrary dense matrices in finite arithmetic [ABB+18, BGVKS23, BGVS24, BGVS22a, BGVS22b].

All existing works examine the setting where i.i.d. noise is added to every entry of a given matrix. This paper asks if it possible to achieve pseudospectral shattering by adding noise to only a subset of entries, selected at random. We provide a positive answer to this question in the regimes when the random perturbation has $O(n \log^2 n)$ and $O(n^{1+\alpha})$ nonzero entries.

Our results are phrased in terms of the sparsity $\rho = \rho(n)$ of the added noise. Given a matrix M, we consider perturbations of the form

$$A = M + N_g, \tag{1}$$

where the entries of N_g are i.i.d. copies of $\delta \cdot g$ where $\delta \sim \text{Bernoulli}(\rho)$ and $g \sim \mathcal{N}(0, 1_{\mathbb{C}})$. As one might expect, our guarantee provides stronger regularization for larger ρ. We measure regularization in terms of the *eigenvector condition number* $\kappa_V(A)$ and *minimum eigenvalue gap* $\eta(A)$. In the following definitions of these quantites, $A = VDV^{-1}$ is any diagonalization of A and $\lambda_1(A), \ldots, \lambda_n(A)$ are the eigenvalues of A, counted with multiplicity:

$$\kappa_V(A) = \inf_{A = VDV^{-1}} \|V^{-1}\| \, \|V\| \quad \text{and} \quad \eta(A) = \min_{i \neq j} |\lambda_i(A) - \lambda_j(A)|.$$

When A is not diagonalizable, we state $\kappa_V(A) = \infty$. Our main result is the following.

Theorem 1.1. *For any $M \in \mathbb{C}^{n \times n}$, if $n\rho = \Omega(\log(n) \log(\|M\| + n))$ then*

$$\Pr\big(\kappa_V(M + N_g) \geq (\|M\| + n^2\rho)^{10K}\big) \leq O\big(n^{-K}\big), \tag{2}$$

and

$$\Pr\big(\eta(M + N_g) \leq (\|M\| + n^2\rho)^{-35K}\big) \leq O\big(n^{-K}\big), \tag{3}$$

where $K = 2\log(n)/\log(n\rho)$.

Thus, we obtain bounds of the form

$$\log \kappa_V(M + N_g) \leq O_\alpha(\log(n)) \quad \text{and} \quad \log(1/\eta(M + N_g)) \leq O_\alpha(\log(n)) \tag{4}$$

with high probability when $\|M\| \leq \text{poly}(n)$ and $\rho(n) = n^\alpha/n$ for any constant $\alpha > 0$. In the sparser regime $\rho(n) = \log^2(n)/n$, we obtain the weaker bounds

$$\log(\kappa_V(M + N_g)) \leq \text{poly}(\log(n)) \quad \text{and} \quad \log(1/\eta(M + N_g)) \leq \text{poly}(\log(n)). \tag{5}$$

Remark 1.2 (Small Perturbations)**.** One can derive a similar bound with smaller perturbations, i.e. σg in place of g, by noting that $\kappa_V(\sigma A) = \kappa_V(A)$ and $\eta(\sigma A) = \sigma\eta(A)$ for any scalar $\sigma > 0$. We find it more convenient to work with $A = M/\sigma + N_g$ instead of $A = M + \sigma N_g$, so that the entries of $A - \mathbb{E}(A)$ have density bounded by 1 everywhere except at 0. Using this observation gives the following Corollary of Theorem 1.1.

Corollary 1.3. *Say $M \in \mathbb{C}^{n \times n}$ has $\|M\| \leq 1$. Fix $\delta > 0$. For sparsity $\rho = \Omega(\frac{\log(n)\log(1/\delta + n)}{n})$, one has*

$$\Pr\left(\kappa_V(M + \delta N_g) \geq \left(\frac{1}{\delta} + n\right)^{10\chi(n)}\right) \leq O\Big(n^{-\chi(n)}\Big), \tag{6}$$

and

$$\Pr\left(\eta(M + \delta N_g) \leq \delta \cdot \left(\frac{1}{\delta} + n\right)^{-35\chi(n)}\right) \leq O\Big(n^{-\chi(n)}\Big), \tag{7}$$

where $\chi(n) = 2\log(n)/\log\log(n)$.

Remark 1.4 (Sharpness of the Bounds)**.** It is easy to see via a coupon collector argument that in the model (1), sparsity $\rho(n) = \Omega(\log(n)/n)$ is required in order to obtain any high probability lower bound on $\eta(A)$. Theorem 1.1 requires a logarithmically denser perturbation than this to obtain (5). We have not made any attempt to optimize the constants in our results in the interest of simplicity.

Algorithmic Significance. There are two ways in which Theorem 1.1 is algorithmically significant. First, it allows the use of sparse rather than dense perturbations in the existing algorithms for diagonalizing dense matrices mentioned above, at the cost of a polylogarithmic increase in the running time. The reason is that in those algorithms, the running time depends polynomially on $\log(\kappa_V(A)/\eta(A))$; this quantity is at most roughly $O(\log(n))$ for dense random perturbations of the input, and at most roughly $O(\log^2(n))$ for sparse perturbations by Theorem 1.1. Passing to sparse perturbations has the effect of reducing the number of random bits required in the execution of the algorithm from roughly $O(n^2 \log n)$ to $O(n \log^3 n)$ (in the regime when the desired backward error is $1/\operatorname{poly}(n)$), since $\log(n)$ bits are used for each nonzero entry of the perturbation. This may be viewed as a step towards derandomizing these algorithms. In the context of numerical analysis, where replicability is desirable, it means that the random seed used by the algorithm can be stored in nearly linear rather than quadratic space.

Second, it may be useful in speeding up iterative algorithms for sparse linear algebra, such as Krylov subspace methods. Iterative algorithms are able to exploit the sparsity of the input matrix by accessing it via matrix vector products. They have running times which generally depend on the number of nonzero entries $\operatorname{nnz}(A)$ of the input matrix A, a quantity which is kept small by a sparse random perturbation. The upshot is that in some cases, the number of iterations depends *logarithmically* on $\kappa_V(A)$, so a sparse random perturbation is able to speed up convergence of the algorithm without blowing up the cost of each iteration. We give a simple example of such an instance in Section 1.3. On the other hand, in contrast to the existing applications of dense pseudospectral shattering, we do not expect iterative algorithms to be able to exploit polynomial lowerbounds on $\eta(A)$ such as those provided by Theorem 1.1. The reason is that all such algorithms which we know of may be viewed as applying a low degree polynomial to the input matrix A, which is not able to resolve polynomially small eigenvalue gaps.

1.1 Related Work

Random Matrix Theory. Several works have studied the distributions of $\sigma_n(\cdot), \kappa_V(\cdot)$, and $\eta(\cdot)$ for $n \times n$ random matrices A where $A - \mathbb{E}(A)$ has i.i.d. entries. The table below lists the relevant works and which quantities they consider, under the normalization $\|\mathbb{E}(A)\| \leq \operatorname{poly}(n)$. These works all take the entries of $N_x = A - \mathbb{E}(A)$ to be of the form $\delta \cdot x$ where $\delta \sim \text{Bernoulli}(\rho)$ is a Bernoulli random variable with parameter $\rho = \rho(n)$, and x is a random variable which varies work to work. From the table, one can see that this

	$\kappa_V(\cdot)$	$\eta(\cdot)$	$\sigma_n(\cdot)$	$\mathbb{E}(A)$	x	$\rho(n)$
[TV07]			✓	any	finite moment conditions	n^α/n
[BR17]			✓	0	finite moment conditions	$\log^c(n)/n$
[RT19]			✓	zI	finite moment conditions	$\omega(1/n)$
[Ge17]		✓	✓	0	finite moment conditions	$\Theta(1)$
[BKMS21]	✓		✓	any	complex Gaussian	1
[BGVKS23]	✓	✓	✓	any	complex Gaussian	1
[BGVKS24, JSS21]	✓	✓	✓	real	bounded real density	1
[EJ24]	✓	✓	✓	real/complex	real/complex Gaussian	1
This work	✓	✓	✓	any	complex Gaussian	$\Omega(\log^2(n)/n)$

work provides the first bounds on $\kappa_V(\cdot)$ and $\eta(\cdot)$ for $\mathbb{E}(A) \neq 0$ and $\rho(n)$ less than 1. In particular, it is the first work on those quantities when $\mathbb{E}(A) \neq 0$ and the entries of $A - \mathbb{E}(A)$ do not have absolutely continuous distributions on either $\mathbb{R}$ or $\mathbb{C}$.

Bounds on $\kappa_V(\cdot)$ and $\eta(\cdot)$. The listed prior works for $\kappa_V(\cdot)$ and $\eta(\cdot)$ obtain bounds of the form $\kappa_V(M + N_x) \leq \operatorname{poly}(n)$ and $\eta(M + N_x) \geq \operatorname{poly}(n)^{-1}$ with high probability for $\|M\| \leq \operatorname{poly}(n)$. The series of works [CEHS24], [CES22], culminating in [EJ24] listed in the table, obtains bounds for real and complex Gaussian perturbations with nearly optimal dependence on n. All of these works bound $\kappa_V(A)$ by obtaining good

control over the size of the ε-*pseudospectrum* of A, defined as

$$\Lambda_\varepsilon(A) = \{z \in \mathbb{C} : \sigma_n(z - A) \le \varepsilon\}.$$

It is well-known that the set $\Lambda_\varepsilon(A)$ always contains disks of radius ε around each eigenvalue of A, and that equality is achieved if and only if $\kappa_V(A) = 1$ [TE05]. A generalization of this fact shows that the limit of $\varepsilon^{-2} \operatorname{vol} \Lambda_\varepsilon(A)$ as $\varepsilon \to 0$ determines $\kappa_V(A)$ up to a poly(n) factor (see [BD20, Corollary 1.5] and [BKMS21, Lemma 3.2]). In order for this limit to be finite, one must show the that $\mathbb{E} \operatorname{vol} \Lambda_\varepsilon(A) = O_n(\varepsilon^2)$ as $\varepsilon \to 0$, which follows from least singular value estimates on shifts of A with the correct exponent, namely:

$$\Pr(\sigma_n(z - A) \le \varepsilon) = O_n(\varepsilon^2) \quad \forall \varepsilon > 0, z \in \mathbb{C}. \tag{8}$$

[BKMS21] prove (8) in the dense complex Gaussian case, yielding a bound on $\kappa_V(A)$; [BGVKS23] then uses this to control $\eta(A)$ as well. In the more challenging setting of real Gaussian perturbations, the bound (8) does not hold for $z \in \mathbb{R}$. The subsequent works [BGVKS24, JSS21, EJ24] overcome this obstacle by first controlling $\eta(A)$, which then allows them to use some alternate arguments which handle $z \in \mathbb{R}$ and $z \in \mathbb{C} \setminus \mathbb{R}$ separately. One such argument is the bootstrapping scheme of [JSS21], which shows that an adequately good lowerbound on $\eta(A)$ and an adequately good upperbound on $\mathbb{E} \operatorname{vol} \Lambda_\varepsilon(A)$ for specific scales $\varepsilon > 0$ (not tending to zero) together imply a bound on $\kappa_V(A)$. While their approach was originally designed to handle the specific difficulties arising in the case of real Gaussian perturbations, we are able to repurpose it sparse complex Gaussian perturbations (see Section 1.2 for details).

We remark that obtaining the exponent of 2 in (8) for complex z (or something close to it) was a sticking point in all previous works. A key technical contribution of this work, described in detail in Section 1.2, is that we show how to obtain bounds on $\kappa_V(A)$ from much weaker least singular value bounds: rather than (8), it suffices to have a bound of type (10) satisfying just (11). Our approach thus reveals that the main bottleneck in proving bounds on $\kappa_V(A)$ is proving bounds on $\eta(A)$. We view this realization as a conceptual contribution of this work.

*Bounds on $\sigma_n(\cdot)$.*The result in the first row of the above table was provided by Tao and Vu in the context of proving the circular law [TV07, Theorem 2.9]. They showed $\sigma_n(M + N) \ge n^{-b}$ with probability at least $1 - n^{-a}$ where b depends on a and α. This estimate was extended to lower sparsity $\rho(n) = \log^c(n)/n$ with the restriction that $\mathbb{E}(A) = 0$ by Basak and Rudelson [BR17], and pushed even further down to any $\rho(n) = \omega(1/n)$ with the milder restriction $\mathbb{E}(A) = zI$ by Rudelson and Tikhomirov [RT19] (though in that case, b must be replaced by $b \log^2(n)$).

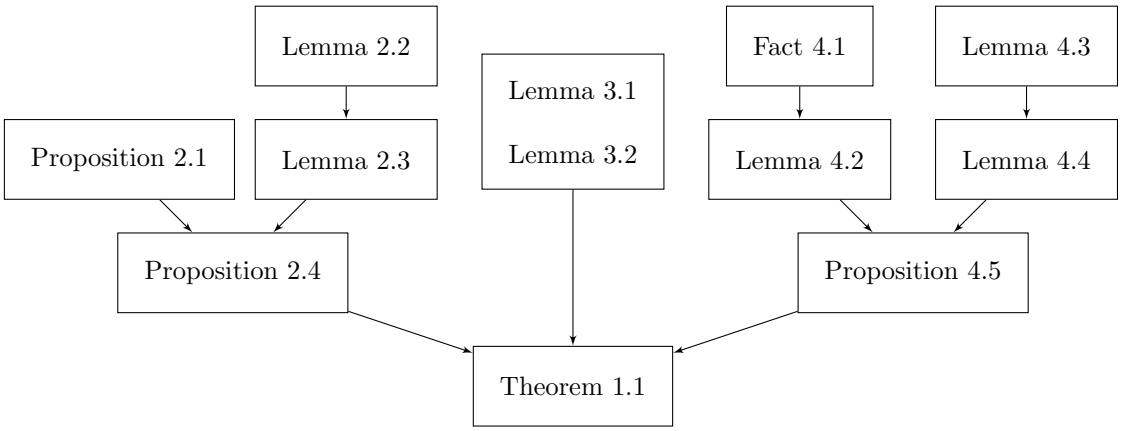
As explained in Section 1.2, bounds on the least singular value play a key role in our results, and we use many ideas from [TV07]. Our technical innovation here is that by restricting attention to complex Gaussian x, we are able to avoid the use of additive combinatorics and prove stronger bounds than previous works, which are needed in this context. In particular, we obtain an improved dependence $b = a/2 + O(1)$ in the Tao-Vu bound. d1.1.

Numerical Analysis. The idea of using a small random perturbation to improve the performance of numerical algorithms (specifically, Gaussian elimination) originated in the influential work [SST06], which analyzed the least singular value under a dense Gaussian perturbation. The first paper to propose the use of a (not necessarily random) perturbation to tame spectral instability in the context of diagonalization algorithms was [Dav07].

[ABB$^+$18] gave the first numerically stable algorithm for diagonalizing a dense non-Hermitian matrix; a key step there was to add a dense complex Gaussian perturbation to the input, and show that it yields a well-conditioned diagonalization problem in a certain geometric sense. The more recent works [BGVKS23, BGVS22a, BGVS22b] used pseudospectral shattering in their analyses of nearly matrix multiplication time and $O(n^3)$ time algorithms for computing the eigendecomposition and Schur form of a matrix. This technique was extended to computing invariant subspaces of Hermitian pencils by [SML24], to diagonalizing arbitrary matrix pencils by [DDS24b], and to computing the Schur form of arbitrary matrix pencils by [DDS24a].

1.2 Technical Overview

We now describe the proof of Theorem 1.1. The proof consists of three steps. In the first two steps, A can be any random matrix. In the third step, we need our particular model of sparse perturbations, $A = M + N_g$. When we omit the subscript we mean $N = N_g$. We dedicate a section to each step. The bulk of the technical work is in the first and third steps; the second step can be viewed as the glue which combines them to produce the final theorem. The logical structure of this paper is captured by the following diagram.



Step 1. Reduction to pseudospectral area and eigenvalue gap (Section 2): We first show (Proposition 2.4) that an upper tail bound on $\Pr(\kappa_V(A) \geq \tau)$ may be derived from a lower tail bound on $\eta(A)$ *and* a bound of type

$$\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A) \leq \mathrm{poly}(n) \cdot \varepsilon^c + \exp(-\log^2(n)/\log(n\rho)), \tag{9}$$

for any constant $c > 0$, where ε is a carefully chosen spectral scale depending c and on the desired deviation τ. The proof is an adaptation of the bootstrapping scheme of [JSS21] and additionally invokes a deterministic exponential upperbound on $\kappa_V(A)$ in terms of $\eta(A)$ (Proposition 2.1). In particular, the allowance of the additive term is essential in the sparse Gaussian case as there is some nontrivial probability that one or more rows and columns of the matrix end up with no noise added to them.

Step 2. Reduction to bottom two singular values (Section 3): The previous step shows we need control over $\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)$ and $\eta(A)$. As mentioned above, $\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)$ reduces to lower tail estimates for the least singular value $\sigma_n(z-A)$ for $z \in \mathbb{C}$ (see Lemma 3.1). Following [BGVKS24, JSS21], a lower tail bound on $\eta(A)$ can be reduced to lower tail bounds on the *bottom two* singular values $\sigma_n(z-A), \sigma_{n-1}(z-A)$ via Weyl majorization and a standard net argument in the complex plane (see Lemma 3.2).

The strength of a tail bound on σ_{n-m} can be characterized in terms of the *power* c_m of ε on the right-hand side of a bound of the form

$$\Pr(\sigma_{n-m}(z-A) \leq \varepsilon) \leq \mathrm{poly}(n)\varepsilon^{c_m} + \beta(n), \tag{10}$$

where $\beta(n)$ is an additive error depending on n. In order to obtain a nontrivial bound on $\eta(A)$, our argument requires (10) to hold for $m = 0, 1$ with

$$\frac{1}{c_0} + \frac{1}{c_1} < 1. \tag{11}$$

Step 3. Control on bottom two singular values (Section 4): We show the required control (10),(11) over the bottom two singular values holds with room to spare. Specifically, we show a bound of the form (10) holds for $c_0 = 2$ and $c_1 = 4$ with $\beta(n) = O(\exp(-\rho n))$. (in fact, we show it holds for $c_m = 2m + 2$ for any constant m, see Proposition 4.5). The argument is based on an ε-net construction following the compressible/incompressible or rich/poor decomposition in [TV07]. That work considers lower tail bounds

of $\sigma_n(M + N_x)$ where x is a general subgaussian random variable and the sparsity parameter $\rho(n) = n^{\alpha-1}$, $\alpha > 0$ is a fixed polynomial in n. They show for every $a > 0$ there exists a $b > 0$ such that

$$\Pr\big(\sigma_n(A) \le n^{-a}\big) \le n^{-b}.$$

By tracing their argument, one can show there is a linear relationship between a and b so that their bound more closely resembles the form (10) for an unspecified tiny constant $c_0 \le b/a$ and $\varepsilon = \frac{1}{\mathrm{poly}(n)}$. This gives enough control over $\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)$ to satisfy (9) in Step 2, but does not satisfy (11) and is therefore not enough to bound $\eta(A)$ (see Remark 1.5).

We resolve this issue by specializing to the complex Gaussian case $x = g$ and achieve three advantages over [TV07] in this setting. Firstly, our argument applies to every m (not just $m = 0$), and we are able to obtain c_0 and c_1 satisfying (11) (as a bonus, our proof yields the optimal power of $c_0 = 2$ for $m = 0$). Secondly, we obtain tail bounds at scales smaller than inverse polynomial in n. Lastly, we are able to push the sparsity parameter down to $O(\log^2(n)/n)$. Furthermore, because g has a continuous density, we avoid the additive combinatorics required by [TV07], resulting in much simpler proofs.

In summary, while our proof uses adaptations of arguments from [JSS21, TV07, BGVKS24] as ingredients, it combines them in a new and more efficient way which clarifies their limits and allows us to handle the sparse setting.

Remark 1.5 (Subgaussian and non-Gaussian perturbations)**.** Much of the proof of Theorem 1.1 goes through for *any* random matrix model, not just $A = M + N_g$. When one replaces complex Gaussian g with a subgaussian random variable, part of the proof of Theorem 1.1 (Lemma 2.4 in particular) can be combined with the existing lower tail bound on $\sigma_n(A)$ from Theorem 2.1 of [TV07] to produce a result of the form

$$\Pr\left(\kappa_V(M + N) \le \mathrm{poly}(n)\right) \ge \Pr\big(\eta(M + N) \ge \mathrm{poly}(n)^{-1}\big) - \mathrm{poly}(n)^{-1}$$

when $\log(n)/\log(n\rho)$ is a constant and $\|M\| \le \mathrm{poly}(n)$. The reason we are not able to fully generalize Theorem 1.1 to the subgaussian setting is that we do not know how to prove strong enough tail lower tail estimates on $\eta(M + N)$.

All of our results hold if complex Gaussians are replaced by any sufficiently concentrated absolutely continuous distribution on $\mathbb{C}$, with bounds depending on the ℓ_∞ norm of the density and the decay of the tails. We have chosen to write the paper in terms of complex Gaussians for ease of exposition.

Notation. $\wedge$ denotes logical "and" (intersection of events); $\vee$ denotes logical "or" (union of events); $\neg$ denotes logical negation (complement of event); $\mathbb{S}^{n-1}$ denotes the complex sphere $\{z \in \mathbb{C}^n : \|z\| = 1\}$; the matrix norms $\|\cdot\| = \|\cdot\|_2, \|\cdot\|_\infty$ are the operator norms. $spr(\cdot)$ is the spectral radius. Matrices are $n \times n$ unless otherwise specified.

1.3 Algorithmic application

As alluded to already, establishing control over the eigenvector condition number of a matrix is essential for rigorous analysis of non-Hermitian eigenvalue problems. The work of [BKMS21] does this by adding a dense perturbation N. The drawback is that the cost of computing matrix-vector products goes from $O(\mathrm{nnz}(M))$ to $O(\mathrm{nnz}(M) + \mathrm{nnz}(N)) = O(n^2)$ where $\mathrm{nnz}(A)$ is the number of nonzero entries in the matrix A. The algorithmic content of this paper is that it suffices to take N to be a sparse perturbation, with $\mathbb{E}\,\mathrm{nnz}(N) = n^2\rho$ for $\rho = \Omega(\log(n)^2/n)$.

As a simple example of an application of Theorem 1.1, we show it implies an algorithm for computing the spectral radius $spr(A)$ of any matrix up to mixed forwards-backwards error running in $\widetilde{O}\big(\mathrm{nnz}(M) + n^2\rho\big)$ time. Computing the spectral radius of a nearby matrix is relevant to the stability analysis of discrete time dynamical systems, see e.g. [Tre05, Chapter IV]; in that context, the algorithm below certifies whether a sparse random perturbation of a given dynamical system stable or unstable upto a desired error.

Our simple algorithm is based on the simple, well-known inequalities:

$$spr(A) \le \|A^k\|^{1/k} = \|VD^kV^{-1}\|^{1/k} \le \|V\|^{1/k}\,\|D\|\,\|V^{-1}\|^{1/k} = \kappa_V(A)^{1/k} spr(A)$$

and the fact that $\|A^k b\|$ is concentrated around $\|A^k\|_F \le \sqrt{n}\|A^k\|$ when b is a Gaussian vector. So when $\kappa_V(A)$ is controlled, we obtain good approximations of the spectral radius. We note that it is unclear how to compute $spr(A)$ efficiently in the absence of a bound on $\kappa_V(A)$.

Theorem 1.6. *Fix any $M \in \mathbb{C}^{n\times n}$ with $\|M\| \ge 1$. For each ρ satisfying $n\rho = \Omega(\log(n)\log(n+\|M\|))$ and $\varepsilon, \delta \in (0,1)$, there is an algorithm $\texttt{specr}(A)$ which computes a value such that*

$$(1-\varepsilon)spr(A+E) \le \texttt{specr}(A) \le (1+\varepsilon)spr(A+E)$$

for some $\|E\| \le \delta$. It uses

$$O\left(\frac{\log(n)}{\log(n\rho)} \cdot \frac{\log(n\|M\|/\delta)}{\varepsilon} \cdot \left(\mathrm{nnz}(M) + n^2\rho\right)\right)$$

operations.

Proof. Sample N with sparsity ρ and set $A = M + (\delta/n)N$. Sample a standard Gaussian vector b. Compute $A^k b$ for k which we specify later in $O(k \cdot \mathrm{nnz}(A))$ time using k matrix-vector multiplies. Then

$$\Pr\left(\frac{1}{n}\left\|A^k\right\|_F \le \left\|A^k b\right\| \le n\left\|A^k\right\|_F\right) \ge 1 - \frac{1}{n}.$$

Conditioned on this event,

$$n^{-1/k} spr(A) \le n^{-1/k}\left\|A^k\right\|_2^{1/k} \le n^{-1/k}\left\|A^k\right\|_F^{1/k} \le \texttt{specr}(A)$$

and

$$\texttt{specr}(A) \le n^{1/k}\left\|A^k\right\|_F^{1/k} \le n^{1.5/k}\left\|A^k\right\|_2^{1/k} \le n^{1.5/k}\kappa_V(A)^{1/k} spr(A).$$

If $k \ge 2\log(n\kappa_V(A))/\varepsilon$, then this gives

$$(1-\varepsilon)spr(A) \le \texttt{specr}(A) \le (1+\varepsilon)spr(A).$$

Note $\kappa_V(A) = \kappa_V((n/\delta)M + N)$ so Theorem 1.1 means with high probability that

$$\log(n\kappa_V(A)) \le \frac{\log(n)}{\log(n\rho)} \cdot (\log(n+\|M\|) + \log(1/\delta)).$$

So one should take

$$k = \Omega\left(\frac{\log(n)}{\log(n\rho)} \cdot \frac{\log(n\|M\|/\delta)}{\varepsilon}\right)$$

to obtain the desired guarantee. □

2 Reduction to eigenvalue spacing and pseudospectral area

Since a matrix with distinct eigenvalues must be diagonalizable, $\eta(A) > 0$ implies $\kappa_V(A) < \infty$. We begin with a quantitative version of this statement: $\kappa_V(A)$ is at most exponential in $1/\eta(A)$.

Proposition 2.1 (Exponential κ_V bound)**.** *Let $\|A\| \le 1$ and $\eta = \eta(A)$. Then*

$$\kappa_V(A) \le n \cdot 2^n \cdot \eta^{1-n}.$$

Proof. The case $\eta = 0$ is trivial so suppose $\eta > 0$. Let $A = VDV^{-1}$ for V with unit column vectors. Number the eigenvalues $\lambda_1, \ldots, \lambda_n$ so that $D_{ii} = \lambda_i$. Let

$$d_J = \inf\{\|Vx\| : \|x\| = 1, \mathrm{supp}(x) \subset J\}$$

so that $d_{[n]} = \sigma_n(V)$ and $\kappa_V(A) \le \sqrt{n}/d_{[n]}$. Let x^J be such that $d_J = \left\|Vx^J\right\|$. We argue via induction that

$$d_J \ge \frac{\eta^{|J|-1}}{2^{|J|}\sqrt{n}}.$$

The base case when J is a singleton follows from the columns of V being unit vectors. Denote $i_J(x) = \arg\min_{j\in J}|x_j|$. Then since $\|A\| \le 1$ we have

$$\begin{aligned} d_J = \left\|Vx^J\right\| \ge \left\|AVx^J\right\| = \left\|VDx^J\right\| &\ge \left\|V(D-\lambda_{i_J(x^J)}I)x^J\right\| - \left|\lambda_{i_J(x^J)}\right| \left\|Vx^J\right\| \\ &\ge \left\|V(D-\lambda_{i_J(x^J)}I)x^J\right\| - d_J. \end{aligned} \tag{12}$$

Consider the vector $(D-\lambda_{i_J(x^J)}I)x^J$. It is supported on $J - \{i_J(x)\}$, and each absolute entry is at least η times the corresponding entry of x^J in absolute value. Thus its magnitude is at least

$$\eta\sqrt{1-\min_{j\in J}|x_j|^2} \ge \eta\sqrt{1-\frac{1}{|J|}}$$

Therefore,

$$2d_J \ge \eta\sqrt{1-\frac{1}{|J|}}\cdot d_{J-i_J(x^J)}$$

which establishes the inductive step. The result then follows from $\prod_{j=2}^{n}(1-1/j) = 1/n$. □

For a more refined result, we define the *eigenvalue condition numbers*. Suppose A has distinct eigenvalues with spectral expansion $A = \sum_j \lambda_j v_j w_j^*$. The eigenvalue condition number of λ_j is defined as

$$\kappa(\lambda_j) = \left\|v_j w_j^*\right\|.$$

Bounds on the eigenvalue condition numbers imply bounds on $\kappa_V(A)$ as

$$\max_j \kappa(\lambda_j) \le \kappa_V(A) \le \sqrt{n\sum_{j=1}^{n}\kappa(\lambda_j)^2} \le n\max_j \kappa(\lambda_j). \tag{13}$$

It is well-known that the eigenvalue condition numbers of a matrix specify the rate at which the ε-pseudospectrum shrinks in the $\varepsilon \to 0$ limit. The following lemma is a non-asymptotic version of this result.

Lemma 2.2 (Disks inside pseudospectrum)**.** *Suppose A has distinct eigenvalues $\lambda_1, \cdots, \lambda_n$ ordered such that $\kappa(\lambda_1) = \max_j \kappa(\lambda_j)$. Then*

$$\Lambda_\varepsilon(A) \supset \mathrm{Ball}\left(\lambda_1, \frac{1}{2}\min\left(\frac{\eta(A)}{n}, \kappa(\lambda_1)\varepsilon\right)\right).$$

Proof. Let $\sum_j \lambda_j v_j w_j^*$ be the spectral expansion of A. Let $r < \eta = \eta(A)$ and $z \in \mathrm{Ball}(\lambda_1, r)$. Then by the triangle inequality,

$$\left\|(zI-A)^{-1}\right\| = \left\|\sum_j \frac{v_j w_j^*}{z-\lambda_j}\right\| \ge \frac{\kappa(\lambda_1)}{r} - \sum_{j\ge 2}\frac{\kappa(\lambda_j)}{\eta - r} \ge \kappa(\lambda_1)\cdot\left(\frac{1}{r} - \frac{n-1}{\eta-r}\right) =: f(r).$$

Observe

$$r(\eta-r)\left(f(r) - \frac{1}{\varepsilon}\right) = \kappa(\lambda_1)\eta - r\cdot\left(n\kappa(\lambda_1) + \frac{\eta}{\varepsilon}\right) + \frac{r^2}{\varepsilon} \ge \kappa(\lambda_1)\eta - r\cdot\left(n\kappa(\lambda_1) + \frac{\eta}{\varepsilon}\right).$$

The right-hand side is nonnegative when

$$r \le \frac{\kappa(\lambda_1)\eta}{\kappa(\lambda_1)n + \eta \cdot \frac{1}{\varepsilon}} = \left(\left(\frac{\eta}{n}\right)^{-1} + (\kappa(\lambda_1)\varepsilon)^{-1}\right)^{-1}$$

On the other hand, nonnegativity implies $f(r) \ge 1/\varepsilon$, which implies $z \in \Lambda_\varepsilon(A)$. Therefore, if

$$r \le \frac{1}{2}\min\left(\frac{\eta}{n}, \kappa(\lambda_1)\varepsilon\right)$$

then $\mathrm{Ball}(\lambda_1, r) \subset \Lambda_\varepsilon(A)$. Lemma 2.2 follows. □

We will use Lemma 2.2 to apply a bootstrapping argument to bound κ_V. The following lemma is a single tug of the bootstrap, and closely follows [JSS21].

Lemma 2.3 (Single bootstrap tug)**.** *Let A be any random matrix and let $F(\tau)$ denote the event that A has distinct eigenvalues $\lambda_1, \dots \lambda_n$ with*

$$\max_j \kappa(\lambda_j)^2 \le \tau.$$

Fix any $\tau_1 < \tau_2$ and $\varepsilon, \eta_0 > 0$. If $\eta_0 \ge n\varepsilon\sqrt{\tau_2}$ then

$$\Pr(F(\tau_1) \wedge \eta(A) \ge \eta_0) \ge \Pr(F(\tau_2) \wedge \eta(A) \ge \eta_0) - \frac{4}{\pi} \cdot \frac{\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)}{\varepsilon^2\tau_1}.$$

Proof. Without loss of generality take $\kappa(\lambda_1) = \max_j \kappa(\lambda_j)$. By Markov's inequality,

$$\Pr(\neg F(\tau_1) \,|\, F(\tau_2) \wedge \eta(A) \ge \eta_0) \le \frac{\mathbb{E}\big(\kappa(\lambda_1)^2 \,|\, F(\tau_2) \wedge \eta(A) \ge \eta_0\big)}{\tau_1}.$$

Conditioned on $F(\tau_2)$, we have $\eta_0 \ge n\varepsilon\sqrt{\tau_2} \ge n\varepsilon\kappa(\lambda_1)$, so by Lemma 2.2 we have

$$\mathrm{vol}\,\Lambda_\varepsilon(A) \ge \frac{\pi}{4}\kappa(\lambda_1)^2\varepsilon^2.$$

Rearranging gives

$$\begin{aligned}\Pr(\neg F(\tau_1) \,|\, F(\tau_2) \wedge \eta(A) \ge \eta_0) &\le \frac{\frac{4}{\pi}\,\mathbb{E}(\mathrm{vol}\,\Lambda_\varepsilon(A) \,|\, F(\tau_2) \wedge \eta(A) \ge \eta_0)}{\varepsilon^2\tau_1} \\ &\le \frac{\frac{4}{\pi}\,\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)}{\varepsilon^2\tau_1 \cdot \Pr(F(\tau_2) \wedge \eta(A) \ge \eta_0)}.\end{aligned} \tag{14}$$

On the other hand,

$$\Pr(F(\tau_1) \,|\, F(\tau_2) \wedge \eta(A) \ge \eta_0) \le \frac{\Pr(F(\tau_1) \wedge \eta(A) \le \eta_0)}{\Pr(F(\tau_2) \wedge \eta(A) \ge \eta_0)}.$$

These two expressions sum to at least 1. Rearranging gives the result. □

Proposition 2.4 (Complete bootstrap)**.** *Let A be any random matrix. Fix any $\tau_0 > 0$ and $R > \eta_0 > 0$. Let $\varepsilon = \eta_0/\left(n\tau_0^{1/2}\right)$. Then*

$$\Pr\left(\max_j \kappa(\lambda_j)^2 \le \tau_0\right) \ge \Pr(\eta(A) \ge \eta_0) - \left[3n \lg \frac{3R}{\eta_0} + 3 \lg \frac{2\eta_0}{\tau_0}\right] \cdot \frac{\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)}{\tau_0 \cdot \varepsilon^2} - \Pr(\|A\| > R). \tag{15}$$

Proof. The idea is to apply Lemma 2.3 with a sequence of thresholds τ. Set

$$\tau_j = \tau_0 \cdot 2^j \quad \text{and} \quad \varepsilon_j = \frac{\eta_0}{n\tau_j{}^{1/2}}.$$

As before, let $F(\tau)$ be the event $\max_j \kappa(\lambda_j)^2 \le \tau$. Let G be the event $\eta(A) \ge \eta_0$. Then Lemma 2.3 implies

$$\Pr(F(\tau_j) \wedge G) \ge \Pr(F(\tau_{j+1}) \wedge G) - \frac{4}{\pi} \cdot \frac{\tau_{j+1}}{\tau_j} \cdot \frac{n^2}{\eta_0^2} \cdot \mathbb{E}\,\mathrm{vol}\,\Lambda_{\varepsilon_{j+1}}(A). \tag{16}$$

Note $\tau_{j+1}/\tau_j = 2$. So by recursively applying Lemma 2.3 for $j = 1, \cdots, K$, we obtain

$$\begin{aligned} \Pr(F(\tau_0) \wedge G) &\ge \Pr(F(\tau_K) \wedge G) - \frac{8}{\pi} \cdot \frac{n^2}{\eta_0^2}\, \mathbb{E} \sum_{j=1}^{K} \mathrm{vol}\,\Lambda_{\varepsilon_j}(A). \\ &\ge \Pr(F(\tau_K) \wedge G) - \frac{8}{\pi} \cdot \frac{n^2}{\eta_0^2} \cdot K \cdot \mathbb{E}\,\mathrm{vol}\,\Lambda_{\varepsilon_0}(A). \end{aligned} \tag{17}$$

Set

$$K = \left\lceil n \lg \frac{3R}{\eta_0} + \lg \frac{\eta_0}{\tau_0} \right\rceil \le n \lg \frac{3R}{\eta_0} + \lg \frac{2\eta_0}{\tau_0}.$$

Then $\tau_K \ge \eta_0(3R/\eta_0)^n$ and so Lemma 2.1 combined with (13) implies $G \wedge \|A\| \le R \implies F(\tau_K)$. Hence

$$\begin{aligned} \Pr(F(\tau_K) \wedge G) &\ge \Pr(F(\tau_K) \wedge G \wedge \|A\| \le R) \\ &= \Pr(F(\tau_K) \,|\, G \wedge \|A\| \le R) \cdot \Pr(G \wedge \|A\| \le R) \\ &= \Pr(G \wedge \|A\| \le R) \\ &\ge \Pr(G) - \Pr(\|A\| > R). \end{aligned} \tag{18}$$

The result follows from combining (17), (18) and writing η_0 in terms of ε_0, τ_0. □

3 Reduction to singular values

In this section we reduce bounds on $\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(A)$ and $\eta(A)$ to estimates on small singular values.

Lemma 3.1 (Pseudospectral area to smallest singular value)**.** *Fix $M \in \mathbb{C}^{n\times n}$ and $R \ge \max(2\|M\|, n)$. Let N have i.i.d. symmetric subgaussian entries. Then*

$$\mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(M+N) \le \pi R^2 \sup_{\|M'\| \le 1.5R+\varepsilon} \Pr(\sigma_n(M'+N) \le \varepsilon) + \exp(-\Omega(R)).$$

Proof. By Fubini,

$$\begin{aligned} (\star) := \mathbb{E}\,\mathrm{vol}\,\Lambda_\varepsilon(M+N) &= \mathbb{E} \int_{\mathbb{C}} \begin{cases} 1 & z \in \Lambda_\varepsilon(M+N) \\ 0 & \text{o.w.} \end{cases} \mathrm{d}z \\ &= \int_{\mathbb{C}} \Pr(\sigma_n(A - zI) \le \varepsilon)\, \mathrm{d}z. \end{aligned} \tag{19}$$

Note $\sigma_n(M+N-zI) \ge |z| - \|M+N\|$. Thus

$$\begin{aligned} (\star) &= \pi R^2 \sup_{|z| \le R+\varepsilon} \Pr(\sigma_n(M+N-zI) \le \varepsilon) + \int_{|z|>R+\varepsilon} \Pr(|z| - \varepsilon < \|A\|) \\ &\le \pi R^2 \sup_{|z| \le R+\varepsilon} \Pr(\sigma_n(M+N-zI) \le \varepsilon) + 2\pi \int_R^\infty x \Pr(x < \|M+N\|)\, \mathrm{d}x \\ &\le \pi R^2 \sup_{\|M'\| \le 1.5R+\varepsilon} \Pr(\sigma_n(M'+N) \le \varepsilon) + 2\pi \int_{R-\|M\|}^\infty 2x \Pr(x < \|N\|)\, \mathrm{d}x. \end{aligned} \tag{20}$$

The final term is bounded by $\exp(-\Omega(R))$ since $\|N\|$ has subgaussian concentration around $\sqrt{n}$, and x is at least $R - \|M\| \geq R/2 \gg \sqrt{n}$. □

The following simple lemma, similar to net arguments appearing in [BGVKS24, JSS21], relates minimum gap bounds to singular value tail bounds.

Lemma 3.2 (Spacing to smallest two singular values)**.** *Let A be any random matrix. Fix any constants $R > r > 0$ and $\lambda \in (0,1)$. Then*

$$\begin{aligned}\Pr(\eta(A) \leq r) \leq \frac{8R^2}{r^2} \sup_{|z|\leq R} \left[\Pr\left(\sigma_n(A - zI) \leq r^{1+\lambda}\right) + \Pr\left(\sigma_{n-1}(A - zI) \leq r^{1-\lambda}\right)\right] \\ + \Pr(spr(M + N) \geq R).\end{aligned} \tag{21}$$

Proof. Let $\mathcal{N} = (r\mathbb{Z} + re^{2\pi i/3}\mathbb{Z}) \cap \mathrm{Ball}(0, R)$. Let λ_i, λ_j be the eigenvalues of A achieving $|\lambda_i - \lambda_j| = \eta(A)$. Consider the event that $\eta(A) \leq r$. If either of λ_i, λ_j are outside $\mathrm{Ball}(0, R)$, then by definition we have $spr(A) \geq R$. On the other hand, if they're both inside the disk, then there must be some $z \in \mathcal{N}$ such that $\lambda_i, \lambda_j \in \mathrm{Ball}(z, r)$. (To see this, note that $\lambda_j \in \mathrm{Ball}(\lambda_i, r) \subset \bigcup_{z\in\mathcal{N}\cap\mathrm{Ball}(\lambda_i,r)} \mathrm{Ball}(z, r)$). So by union bound,

$$\Pr(\eta(A) \leq r) \leq \Pr(spr(A) \geq R) + \sum_{z\in\mathcal{N}} \Pr\left(|\lambda_i - z||\lambda_j - z| \leq r^2\right).$$

Note that $\lambda_i - z$ and $\lambda_j - z$ are eigenvalues of $A - zI$, so by Weyl majorization (see [HJ91, Theorem 3.3.14]) we deterministically have

$$|\lambda_i - z||\lambda_j - z| \geq \sigma_n(A - zI)\sigma_{n-1}(A - zI). \tag{22}$$

Finally, the quantity (22) being upper bounded by r^2 implies one of the following events must occur:

$$\sigma_n(A - zI) \leq r^{1+\lambda} \ \vee \ \sigma_{n-1}(A - zI) \leq r^{1-\lambda}.$$

Therefore, for each $z \in \mathcal{N}$ we have by union bound

$$\Pr\left(|\lambda_i - z||\lambda_j - z| \leq r^2\right) \leq \Pr(\sigma_n(A - zI) \leq r^{1+\lambda}) + \Pr(\sigma_{n-1}(A - zI) \leq r^{1-\lambda}) \tag{23}$$

The result follows by noting $|\mathcal{N}| \leq (4\pi/\sqrt{3})(R/r)^2$. □

We note that given bounds of type (10), the conclusion of Lemma 3.2 is nontrivial whenever

$$\sup_{\lambda\in(0,1)} \min((1 + \lambda)c_0, (1 - \lambda)c_1) > 2. \tag{24}$$

The left hand side is optimized for $\lambda = (c_1 - c_0)/(c_1 + c_0)$ in which case (24) reduces to (11).

4 Bounds on singular values

Section 3 showed that control over $\mathbb{E}\,\Lambda_\varepsilon(A)$ and $\eta(A)$ can be reduced to control over the bottom two singular values $\sigma_n(z - A)$ and $\sigma_{n-1}(z - A)$. As described in the introduction, this section shows tail bounds of the form (10) for $m = 0, 1$ with $c_0 = 2$ and $c_1 = 4$.

This is the only section where we strongly use the model of random matrix $A = M + N_g$. In particular, it is important that the entries of N_g are independent and are of the form $\delta \cdot g$ where δ is Bernoulli and g has an absolutely continuous distribution on $\mathbb{C}$. We introduce some helpful notation for this section. The Lévy concentration (or small-ball probability) of a random variable X is the function

$$\mathcal{L}(X, r) = \sup_{z\in\mathbb{C}} \Pr(X \in \mathrm{Ball}(z, r)).$$

Let $\vec{g}$ and $\vec{\delta}_\rho$ denote a vector of i.i.d. complex Gaussians and Bernoulli random variables with mean ρ respectively. For any complex vector v we write

$$p(v,r) := \mathcal{L}(\langle \vec{g} \odot \vec{\delta}_\rho, v\rangle, r),$$

where $\odot$ is the Hadamard (entrywise) product. This definition is motivated by the simple observation that the independence of the rows of N_g implies

$$\Pr\big(\|(M+N_g)v\|_\infty \le r\big) \le p(v,r)^n.$$

Partition the sphere based on $p(v,r)$:

$$\texttt{Incomp}(r,s) = \big\{v \in \mathbb{S}^{n-1} : p(v,r) \le s\big\}, \quad \texttt{Comp}(r,s) = \big\{v \in \mathbb{S}^{n-1} : p(v,r) \ge s\big\}.$$

We follow the standard approach for showing a singular value bound. In particular, we show its unlikely for any vector in `Incomp` or in `Comp` to witness the small singular space with different arguments. We begin with the `Incomp` case. The first technical lemma establishes a basis for the small singular space that has a particular form. The strong anti-concentration $\langle \vec{g} \odot \vec{\delta}_\rho, v\rangle$ for $v \in \texttt{Incomp}$ gives the result.

For the `Comp` case, we no longer have strong anti-concentration of $\langle \vec{g} \odot \vec{\delta}_\rho, v\rangle$ so we use the fact that `Comp` has small entropy and construct the appropriate net to union bound over.

Fact 4.1 (Reduced row echelon form basis)**.** *Every subspace of dimension k can be expressed as the column space of a matrix of the form $P\begin{bmatrix} D \\ X \end{bmatrix}$ where P is a permutation matrix, D is diagonal with absolute entries at least $1/\sqrt{n}$, and X is such that the columns of $\begin{bmatrix} D \\ X \end{bmatrix}$ are unit length.*

This fact is simply the result of running Gaussian elimination with complete pivoting on a basis and renormalizing.

Lemma 4.2 (Incompressible vectors don't witness the small singular space)**.** *Fix any $\varepsilon, s \in (0,1)$ and $m \in \mathbb{Z}_{\ge 0}$. Then*

$$\Pr\big(\sigma_{n-m}(M+N_g) \le \varepsilon \wedge \exists v \in \texttt{Incomp}(2\varepsilon\sqrt{n}, s) \text{ s.t. } \|(M+N_g)v\| \le \varepsilon\big) \le \binom{n}{m+1} s^{m+1}.$$

Proof. Suppose $\sigma_{n-m}(M+N_g) \le \varepsilon$. Then there exists a subspace W of dimension $m+1$ such that $\|w^*(M+N_g)\| \le \varepsilon$ for each $w \in W \cap \mathbb{S}^{n-1}$. By Fact 4.1, this means there's a permutation P, diagonal D with $|D_{jj}| \ge 1/\sqrt{n}$, and some X such that

$$\max_j \left\| e_j^* \begin{bmatrix} D \\ X \end{bmatrix}^* P^*(M+N_g) \right\| \le \varepsilon.$$

Condition on a particular value of P. Break into blocks $P^*(M+N_g) = \begin{bmatrix} B_1 \\ B_2 \end{bmatrix}$ where B_1 has $m+1$ rows and B_2 has $n-m-1$ rows. If there exists unit v satisfying $p(v, 2\varepsilon\sqrt{n}) \le s$ and

$$\|(M+N_g)v\| = \sqrt{\|B_1 v\|^2 + \|B_2 v\|^2} \le \varepsilon$$

then there certainly exists unit u satisfying $p(u, 2\varepsilon\sqrt{n}) \le s$ and

$$\|B_2 u\| \le \varepsilon$$

which depends only on B_2. Then

$$\begin{aligned}\varepsilon \geq \max_j \left\| e_j^* \begin{bmatrix} D \\ X \end{bmatrix}^* P^*(M+N_g) \right\| &= \max_j \left\| e_j^*(D^* B_1 + X^* B_2) \right\| \\ &\geq \max_j \left\| e_j^* D^* B_1 u + e_j^* X^* B_2 u \right\| \\ &\geq \max_j \left\| D_{jj} e_j^* B_1 u + e_j^* X^* B_2 u \right\| \end{aligned} \tag{25}$$

which by $\|B_2 u\| \leq \varepsilon, \left\|e_j^* X\right\| \leq 1, |D_{jj}| \geq 1/\sqrt{n}$ implies

$$2\varepsilon\sqrt{n} \geq \max_j \|e_j B_1 u\| = \|B_1 u\|_\infty . \tag{26}$$

Note that u is independent of B_1, so the probability this occurs is $p(u, 2\varepsilon\sqrt{n})^{m+1} \leq s^{m+1}$. Note that when we conditioned on P, we only used which indices corresponded to blocks B_1 and B_2. So we must pay a union bound cost of $\binom{n}{m+1}$. We therefore obtain the overall bound of $\binom{n}{m+1} s^{m+1}$. □

Lemma 4.3 (Compressible vectors are close to sparse)**.** *If $p(v, r) \geq s$, then*

$$\frac{\log(2/s)}{\rho} \geq \#\Big\{ j : |v_j| \geq r\sqrt{2/s} \Big\}.$$

Proof. By iterated expectation we have

$$\begin{aligned} s \leq p(v,r) &= \Pr\Big(\Big| \Big\langle \vec{g} \odot \vec{\delta}, v \Big\rangle \Big| \leq r \Big) \\ &= \mathop{\mathbb{E}}_{\delta} \Pr_g \Big(\Big| \Big\langle \vec{g}, \vec{\delta} \odot v \Big\rangle \Big| \leq r \Big) \\ &\leq \mathop{\mathbb{E}}_{\delta} \min\left(1, \frac{r^2}{\|\vec{\delta} \odot v\|^2} \right). \\ &\leq \Pr\left(\|\vec{\delta} \odot v\|^2 \leq \frac{2r^2}{s} \right) + \frac{s}{2}. \\ &\leq (1-\rho)^{\#\{j : |v_j|^2 \geq 2r^2/s\}} + \frac{s}{2}. \end{aligned} \tag{27}$$

Using $1 - \rho \leq e^{-\rho}$ and rearranging gives the result. □

Lemma 4.4 (Compressible vectors do not witness the small singular space)**.** *Fix $\varepsilon, R > 0$ and integer $K \geq 2$ satisfying*

$$\frac{2^{3(K+1)} \varepsilon^2 n^2 R^{2K}}{\rho^K} \leq s \leq \frac{\rho}{n} \quad \text{and} \quad 2\log\Big(\frac{s}{2\varepsilon^2}\Big) \log(2/s)^{\frac{1}{K+1}} \cdot \frac{n^{\frac{1}{K+1}}}{n} \leq \rho.$$

Then

$$\Pr\big(\exists v \in \mathtt{Comp}(2\varepsilon\sqrt{n}, s) \text{ s.t. } \|(M+N_g)v\| \leq \varepsilon\big) \leq e^{-\frac{1}{2}n\rho} + \Pr\big(\|M+N_g\|_\infty \geq R\big). \tag{28}$$

Proof. Suppose such a v exists. We consider a range of thresholds from $2\varepsilon\sqrt{n} \cdot \sqrt{2/s}$ to $1/\sqrt{n}$ and estimate how many entries of v are above each threshold. The key parameter is the number of thresholds we consider. In this proof we consider $K+1$ exponentially spaced thresholds. Let

$$t = (2\varepsilon\sqrt{n} \cdot \sqrt{2/s} \cdot \sqrt{n})^{1/K}$$

and set

$$a_j(v) = \#\left\{ i : |v_i| \geq \frac{1}{\sqrt{n}} \cdot t^j \right\}$$

to be the number entries of v above the jth threshold. Lemma 4.3 immediately implies that $a_K \leq \frac{\log(2/s)}{\rho}$. On the other hand, $a_0(v) \geq 1$ since v is a unit vector. So

$$\frac{a_0(v)}{a_1(v)} \cdot \frac{a_1(v)}{a_2(v)} \cdots \frac{a_{K-1}(v)}{a_K(v)} = \frac{a_0(v)}{a_K(v)} \geq \frac{\rho}{\log(2/s)}.$$

In particular, there exists an index $k < K$ such that

$$\frac{a_k(v)}{a_{k+1}(v)} \geq \left(\frac{\rho}{\log(2/s)}\right)^{1/K}.$$

From v, produce $\tilde{v}$ by doing the following. Round the entries of v which are less than $\frac{1}{\sqrt{n}} \cdot t^{k+1}$ down to 0. Round the real and imaginary parts of all other entries *up* to the nearest multiple of $\frac{1}{2\sqrt{n}} \cdot t^{k+1}$. Denote the set of all possible such $\tilde{v}$ as $\mathcal{N}_k$. Then

$$\|v - \tilde{v}\|_\infty \leq \frac{1}{\sqrt{n}} \cdot t^{k+1} \quad \text{and} \quad a_k(v) = a_k(\tilde{v}).$$

Thus

$$\begin{aligned} \|(M+N)v\| \leq \varepsilon &\implies \|(M+N_g)v\|_\infty \leq \varepsilon \\ &\implies \|(M+N_g)\tilde{v}\|_\infty \leq \varepsilon + \|M+N_g\|_\infty \|v-\tilde{v}\|_\infty \\ &\implies \|(M+N_g)\tilde{v}\|_\infty \leq \varepsilon + \frac{R}{\sqrt{n}} \cdot t^{k+1} \quad \text{or} \quad \|M+N_g\|_\infty \geq R. \end{aligned} \tag{29}$$

Further partition

$$\mathcal{N}_k = \bigcup_j \mathcal{N}_{k,j}, \quad \mathcal{N}_{k,j} = \mathcal{N}_k \cap \{\tilde{v} : a_k(\tilde{v}) = j\}, \quad 1 \leq j \leq \frac{\log(2/s)}{\rho}.$$

Then for $\tilde{v} \in \mathcal{N}_{k,j}$ we have

$$\begin{aligned} \Pr\left(\|(M+N_g)\tilde{v}\|_\infty \leq \varepsilon + \frac{R}{\sqrt{n}} \cdot t^{k+1}\right) &\leq p\left(\tilde{v}, \varepsilon + \frac{R}{\sqrt{n}} \cdot t^{k+1}\right)^n \\ &\leq \left[(1-\rho)^j + \left(\frac{\varepsilon + \frac{R}{\sqrt{n}} \cdot t^{k+1}}{\frac{1}{\sqrt{n}} t^k}\right)^2\right]^n \\ &= \left[(1-\rho)^j + \left(\sqrt{n}\frac{\varepsilon}{t^k} + Rt\right)^2\right]^n \\ &\leq \left[(1-\rho)^j + \frac{\rho}{2}\right]^n \end{aligned} \tag{30}$$

where the last step follows from the hypotheses of this lemma. We finally apply the estimate

$$\begin{aligned} \left[(1-\rho)^j + \frac{\rho}{2}\right]^n &\leq \max\left((1-\rho)^{j-1}, \frac{1}{2}\right)^n \\ &\leq \exp(-n\min(\rho \cdot (j-1), \log(2))). \end{aligned} \tag{31}$$

Let $\tilde{j} = j\Big(\frac{\log(2/s)}{\rho}\Big)^{1/K}$. Then we have

$$\begin{aligned}|\mathcal{N}_{k,j}| &\le \binom{n}{\tilde{j}}\left(\frac{2\sqrt{n}}{t^{k+1}}\right)^{2\tilde{j}} \\ &\le \binom{n}{\tilde{j}}\left(\frac{1}{\varepsilon\cdot\sqrt{2/s}\cdot\sqrt{n}}\right)^{2\tilde{j}} \\ &\le \exp\left(\log\Big(\frac{es}{2\varepsilon^2}\Big)\cdot j\cdot\left(\frac{\log(2/s)}{\rho}\right)^{1/K}\right).\end{aligned} \tag{32}$$

By combining this with (31), applying union bound over k, j, and using the second hypothesis of the lemma we obtain the desired result. □

Combining Lemma 4.2 and Lemma 4.4 a immediately implies the following.

Proposition 4.5. *Fix $K = \frac{2\log(n)}{\log(n\rho)}$. Then for every $R > 0$ and*

$$0 < \varepsilon \le \frac{\rho^{1+K}}{8n^{4+K}R^{2K}}$$

we have

$$\Pr(\sigma_{n-m}(M+N_g) \le \varepsilon) \le n^{2+m+1}R^{2K(m+1)}\cdot\varepsilon^{2(m+1)} + \exp(-\rho n) + \Pr\big(\|M+N_g\|_\infty \ge R\big).$$

Proof. Because $\texttt{Comp}(2\varepsilon\sqrt{n}, s) \sqcup \texttt{Incomp}(2\varepsilon\sqrt{n}, s)$ is a partitioning the sphere, the event $\sigma_{n-m}(M+N_g) \le \varepsilon$ is the union of the events whose probabilities are bounded in each of Lemma 4.2 and Lemma 4.4. We obtain the final result by a union bound. □

Remark 4.6. For $m > 0$, there is overlap between the two events we union over in Proposition 4.5 and therefore slack in the bound. In particular, when $\sigma_{n-m}(M+N_g)$ is small, there should be an entire $(m+1)$-dimensional space witnessing the small singular space. Lemma 4.2 bounds the probability that $\sigma_{n-m}(M+N_g)$ is small and *a single* vector in `Incomp` witnesses the small singular space. Lemma 4.4 bounds the probability *a single* vector in `Comp` witnesses the small singular space. But it is possible that the $(m+1)$-dimensional small singular space contains both vectors in `Incomp` and `Comp`.

5 Proof of Theorem 1.1

We now prove the main theorem of this paper.

Proof of Theorem 1.1. Let $f(m,\varepsilon) = \sup_{\|M'\|\le R}\Pr(\sigma_{n-m}(M'+N) \le \varepsilon)$. Pick $R \ge \|M\| + n^2\rho$ so that $\Pr(\|M+N\| \le R) \le e^{-n}$. Then combining the reductions Lemma 3.1 and 3.2 with 2.4 gives

$$\Pr\big(\kappa(\lambda_1)^2 \ge \tau\big) \le R^2\left(\frac{f(0,\eta^{4/3}) + f(1,\eta^{2/3})}{\eta^2}\right) + 2n^2\log(R/\eta_0)\frac{f\Big(0,\frac{\eta}{n\sqrt{\tau}}\Big)}{\eta^2} + e^{-n}.$$

Note the first term on the right hand side is $\Pr(\eta(M+N) \ge \eta)$. Set $\tau = \eta^{-2/3}/n^2$. Then

$$\begin{aligned}\Pr\left(\kappa(\lambda_1)^2 \ge \frac{1}{n^2\eta^{2/3}}\right) &\le R^2\left(\frac{f(0,\eta^{4/3}) + f(1,\eta^{2/3})}{\eta^2}\right) + 2n^2\log(R/\eta_0)\frac{f\big(0,\eta^{4/3}\big)}{\eta^2} + e^{-n} \\ &\le (R^2 + 2n^2\log(R/\eta))\frac{f(0,\eta^{4/3})}{\eta^2} + R^2\frac{f(1,\eta^{2/3})}{\eta^2} + e^{-n}\end{aligned} \tag{33}$$

Now we would like to apply Proposition 4.5 with $m = 0$ and $m = 1$ to bound $f(m, \varepsilon)$. The hypothesis of that proposition is satisfied when

$$\eta^{2/3} \leq \frac{\rho^{1+K}}{8n^{4+K}R^{2K}}.$$

Take it to be that upper bound divided by R^{2K+2}. Then Proposition 4.5 gives

$$\begin{aligned}
\Pr\big(\kappa(\lambda_1)^2 \geq 8n^{3+2K}R^{2K}\big) &\leq (R^2 + 2n^2\log(R/\eta))\frac{f(0,\eta^{4/3})}{\eta^2} + R^2\frac{f(1,\eta^{2/3})}{\eta^2} \\
&\leq R^5\frac{f(0,\eta^{4/3})}{\eta^2} + R^2\frac{f(1,\eta^{2/3})}{\eta^2} + e^{-n} \\
&\leq \frac{1}{n^K} + R^{20K}\cdot e^{-\rho n}.
\end{aligned} \tag{34}$$

The assumption that $\rho n = \Omega(\log(n)\log(\|M\| + n))$ guarantees that the second term is less than the first. Similarly we have

$$\begin{aligned}
\Pr(\eta(M+N) \leq \eta) &\leq R^2\frac{f(0,\eta^{4/3})}{\eta^2} + R^2\frac{f(1,\eta^{2/3})}{\eta^2}. \\
&\leq \frac{1}{n^K} + R^{20K}\cdot e^{-\rho n}.
\end{aligned} \tag{35}$$

□

References

[ABB+18] Diego Armentano, Carlos Beltrán, Peter Bürgisser, Felipe Cucker, and Michael Shub. A stable, polynomial-time algorithm for the eigenpair problem. *Journal of the European Mathematical Society*, 20(6):1375–1437, 2018.

[BD20] P. Bourgade and G. Dubach. The distribution of overlaps between eigenvectors of Ginibre matrices. *Probability Theory and Related Fields*, pages 397–464, 2020.

[BGVKS23] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Pseudospectral shattering, the sign function, and diagonalization in nearly matrix multiplication time. *Foundations of computational mathematics*, 23(6):1959–2047, 2023.

[BGVKS24] Jess Banks, Jorge Garza-Vargas, Archit Kulkarni, and Nikhil Srivastava. Overlaps, eigenvalue gaps, and pseudospectrum under real Ginibre and absolutely continuous perturbations. *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 60(4), November 2024.

[BGVS22a] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global Convergence of Hessenberg Shifted QR II: Numerical Stability. *arXiv preprint arXiv:2205.06810*, 2022.

[BGVS22b] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global convergence of Hessenberg shifted QR III: Approximate Ritz values via shifted inverse iteration. *arXiv preprint arXiv:2205.06804*, 2022.

[BGVS24] Jess Banks, Jorge Garza-Vargas, and Nikhil Srivastava. Global convergence of Hessenberg shifted QR I: Exact arithmetic. *Foundations of Computational Mathematics*, pages 1–34, 2024.

[BKMS21] Jess Banks, Archit Kulkarni, Satyaki Mukherjee, and Nikhil Srivastava. Gaussian Regularization of the Pseudospectrum and Davies' Conjecture. *Communications on Pure and Applied Mathematics*, 74(10):2114–2131, 2021.

[BR17] Anirban Basak and Mark Rudelson. Invertibility of Sparse non-Hermitian matrices. *Advances in Mathematics*, 320:426–483, 2017.

[CEHS24] Giorgio Cipolloni, László Erdős, Joscha Henheik, and Dominik Schröder. Optimal lower bound on eigenvector overlaps for non-hermitian random matrices. *Journal of Functional Analysis*, 287(4):110495, 2024.

[CES22] Giorgio Cipolloni, László Erdős, and Dominik Schröder. On the condition number of the shifted real ginibre ensemble. *SIAM Journal on Matrix Analysis and Applications*, 43(3):1469–1487, 2022.

[Dav07] E. Brian Davies. Approximate Diagonalization. *SIAM Journal on Matrix Analysis and Applications*, 29(4):1051–1064, 2007.

[DDS24a] James Demmel, Ioana Dumitriu, and Ryan Schneider. Fast and Inverse-Free Algorithms for Deflating Subspaces, 2024.

[DDS24b] James Demmel, Ioana Dumitriu, and Ryan Schneider. Generalized Pseudospectral Shattering and Inverse-Free Matrix Pencil Diagonalization, 2024.

[EJ24] László Erdős and Hong Chang Ji. Wegner estimate and upper bound on the eigenvalue condition number of non-Hermitian random matrices. *Communications on Pure and Applied Mathematics*, 2024.

[Ge17] Stephen Cong Ge. *The Eigenvalue Spacing of IID Random Matrices and Related Least Singular Value Results*. PhD thesis, University of California, Los Angeles, 2017.

[HJ91] Roger A. Horn and Charles R. Johnson. *Singular value inequalities*, page 177. Cambridge University Press, 1991.

[JSS21] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. On the real Davies' conjecture. *The Annals of Probability*, 49(6):3011 – 3031, 2021.

[RT19] Mark Rudelson and Konstantin Tikhomirov. The sparse circular law under minimal assumptions. *Geometric and Functional Analysis*, 29:561–637, 2019.

[SML24] Aleksandros Sobczyk, Marko Mladenović, and Mathieu Luisier. Invariant subspaces and PCA in nearly matrix multiplication time, 2024.

[SST06] Arvind Sankar, Daniel A Spielman, and Shang-Hua Teng. Smoothed analysis of the condition numbers and growth factors of matrices. *SIAM Journal on Matrix Analysis and Applications*, 28(2):446–476, 2006.

[TE05] Lloyd N Trefethen and Mark Embree. *Spectra and pseudospectra*. Princeton University Press, Princeton, NJ, July 2005.

[Tre05] Lloyd N Trefethen. Spectra and Pseudospectra: The Behaviour of Non-normal Matrices and Operators. In *The graduate student's guide to numerical analysis' 98: Lecture notes from the VIII EPSRC Summer School in Numerical Analysis*, pages 217–250. Springer, 2005.

[TV07] Terence Tao and Van Vu. Random matrices: The circular law. *Communications in Contemporary Mathematics*, pages 261–307, 2007.