Sparse Pseudospectral Shattering

<u>Rikhav Shah</u>, Edward Zeng, Nikhil Srivastava

Abstract

A central question in numerical analysis is the following: how do the eigenvalues and eigenvectors of a matrix behave under perturbations of its entries? For Hermitian 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 non-Hermitian 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. 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 [BVKS20, JSS20], and is currently an essential mechanism in all of the known convergence results about diagonalizing arbitrary dense matrices. Crucially, however, all current work examines the setting where i.i.d. noise is added to every single 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.

In fact, we show only $O(n \log^2(n))$ entries need to be perturbed to achieve sufficient regularization for many downstream algorithmic tasks. Our results are phrased in terms of the sparsity $\rho = \rho(n)$ of the added noise. In our model, each entry of a given matrix M is perturbed by a complex Gaussian g with probability ρ , and left unchanged otherwise. As one might expect, our guarantee provides stronger regularization the larger ρ is. 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.

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

Given a matrix M, the perturbation described above has the form $M + N_g$, where the entries of N_g are i.i.d. copies of $\delta \cdot g$ where $\delta \sim \text{Bernoulli}(\rho)$ and $g \sim \mathcal{N}(0, 1_{\mathbb{C}})$. Our main theorem is as follows.

Theorem 1. Set $K = 2\log(n)/\log(n\rho)$. For any $M \in \mathbb{C}^{n \times n}$, if $n\rho = \Omega(\log(n)\log(||M|| + n))$ then

$$\Pr(\kappa_V(M+N_g) \ge (\|M\| + n^2 \rho)^{10K}) \le O(n^{-K}),$$
(1)

and

$$\Pr(\eta(M+N_g) \le (\|M\|+n^2\rho)^{-35K}) \le O(n^{-K}).$$
(2)

The proof of this theorem 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, $A = M + N_q$.

1. Bootstrapping: An important object related to spectral stability is the ε -pseudospectrum of A, defined as

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

It always contains disks of radius ε around each eigenvalue; equality is achieved if and only if A is a normal matrix, i.e. $\kappa_V(A) = 1$. For less well conditioned matrices, the ε -pseudospectrum will be larger. A quantifiable version of this statement relates the area of the pseudospectrum to both $\kappa_V(\cdot)$ and $\eta(\cdot)$. The bootstrapping argument of [JSS20] turns this observation into a probabilistic tail bound: a strong upper bound on $\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A)$ and a lower tail bound on $\eta(A)$ establishes an upper tail bound on $\kappa_V(A)$. We adapt their argument and improve it by dramatically lessening the control on $\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A)$ required for the argument to go through. The ideal control would be of the form

$$\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A) \le \operatorname{poly}(n) \cdot \varepsilon^2.$$

The bootstrapping argument of [JSS20] shows that it suffices to have $\varepsilon^2 \log(1/\varepsilon)$ in place of ε^2 . We show it suffices to have $\varepsilon^c + \exp(-n)$ for any constant c > 0 in place of ε^2 .

2. Reduction to bottom two singular values: This step relies on known arguments which were also used in [BKMS21, BGVKS23]. The previous step shows we need control over $\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A)$ and $\eta(A)$. As may be clear from the definition, $\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A)$ is immediately convertible to lower tail estimates for the least singular value $\sigma_n(z-A)$ for $z \in \mathbb{C}$. We also show a lower tail bound on $\eta(A)$ can be reduced to lower tail bounds on the *bottom two* singular values $\sigma_n(z-A)$, $\sigma_{n-1}(z-A)$.

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

$$\Pr(\sigma_{n-m}(A) \le \varepsilon) \le \operatorname{poly}(n)\varepsilon^{c_m} + \exp(-n).$$
(3)

The reduction from $\eta(A)$ described in Lemma ?? goes through when

$$\frac{1}{c_0} + \frac{1}{c_1} < 1$$

For sufficient control on $\mathbb{E} \operatorname{vol} \Lambda_{\varepsilon}(A)$, we just need $c_0 > 0$. Thus the bottleneck is the reduction from $\eta(A)$.

3. Control on bottom two singular values: We show the required control over the bottom two singular values holds with room to spare. Specifically, we show a bound of the form (3) holds for $c_0 = 2$ and $c_1 = 4$ (in fact, we show it holds for $c_m = 2m + 2$ for any constant m). 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 sub-Gaussian random variable and the sparsity parameter $\rho(n) = n^{\alpha-1}$, $\alpha > 0$ is a fixed polynomial in n. They show for every polynomial n^A , there exists a polynomial n^B such that

$$\Pr(\sigma_n(A) \le n^{-A}) \le n^{-B}$$

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 (3) for an unspecified tiny constant c_0 and $\varepsilon = \frac{1}{\operatorname{poly}(n)}$.

By our improvement to the bootstrapping argument, their result gives enough control over \mathbb{E} vol $\Lambda_{\varepsilon}(A)$. However, it is not enough for $\eta(A)$. We specialize to the complex Gaussian case x = g (or really the case of x having bounded density on \mathbb{C}) and achieve three advantages over [TV07] in this setting. Firstly, our argument applies to every m (not just m = 0), and we show the optimal power of $c_0 = 2$ in the m = 0 case. Secondly, we may take ε to be arbitrarily small. Lastly, we are able to push the sparsity parameter down to $(\log n)^2/n$.

Furthermore, because g has a continuous density, we avoid the additive combinatorics required by [TV07], resulting in much simpler proofs.

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 $O(\operatorname{nnz}(M))$ to $O(\operatorname{nnz}(M) + \operatorname{nnz}(N)) = O(n^2)$ where $\operatorname{nnz}(A)$ 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 $\mathbb{E}\operatorname{nnz}(N) = n^2\rho$ for $\rho = \Omega(\log(n)^2/n)$.

As a simple example of an application of Theorem 1, we show it implies an algorithm for computing the spectral radius spr(M) of any matrix up to mixed forwards-backwards error ε using just

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

floating point operations.

A full preprint can be found at https://math.berkeley.edu/~rdshah/files/sparsepseudospectralshattering.pdf

References

- [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.
- [BVKS20] Jess Banks, Jorge Garza Vargas, Archit Kulkarni, and Nikhil Srivastava. Overlaps, eigenvalue gaps, and pseudospectrum under real ginibre and absolutely continuous perturbations, 2020.

[JSS20] Vishesh Jain, Ashwin Sah, and Mehtaab Sawhney. On the Real Davies' Conjecture, 2020.

- [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.
- [TV07] Terence Tao and Van Vu. Random matrices: The circular law. Communications in Contemporary Mathematics, pages 261–307, 2007.
- [BR17] Anirban Basak and Mark Rudelson. Invertibility of sparse non-hermitian matrices. Advances in Mathematics, 320:426–483, 2017.