Version of Tuesday 27th March, 2018, 09:47

# COALESCING POINTS FOR EIGENVALUES OF BANDED MATRICES DEPENDING ON PARAMETERS WITH APPLICATION TO BANDED RANDOM MATRIX FUNCTIONS

LUCA DIECI, ALESSANDRA PAPINI, AND ALESSANDRO PUGLIESE

Abstract. In this work, we develop and implement new numerical methods to locate generic degeneracies (i.e., isolated parameters' values where the eigenvalues coalesce) of banded matrix valued functions. More precisely, our specific interest is in two classes of problems: (i) symmetric, banded, functions $A(x) \in \mathbb{R}^{n\times n}$, smoothly depending on parameters $x \in \Omega \subset \mathbb{R}^2$, and (ii) Hermitian, banded, functions $A(x) \in \mathbb{C}^{n\times n}$, smoothly depending on parameters $x \in \Omega \subset \mathbb{R}^3$.

The computational task of detecting coalescing points of banded parameter dependent matrices is very delicate and challenging, and cannot be handled using existing eigenvalues' continuation approaches. For this reason, we present and justify new techniques that will enable continuing path of eigendecompositions and reliably decide whether or not eigenvalues coalesce, well beyond our ability to numerically distinguish close eigenvalues.

As important motivation, and illustration, of our methods, we perform a computational study of the density of coalescing points for random ensembles of banded matrices depending on parameters. Relatively to random matrix models from truncated GOE and GUE ensembles, we will give computational evidence in support of power laws for coalescing points, expressed in terms of the size and bandwidth of the matrices.

## 1. Introduction

In this work, we consider families of banded matrices smoothly depending on parameters, and we are interested in locating parameter values where the eigenvalues coalesce. To be precise, we will consider banded symmetric (real valued) functions depending on two real parameters, and banded Hermitian (complex valued) functions depending on three real parameters. The bandwidth, hereafter indicated with $b$, will always be $b \geq 1$, the case $b = 1$ corresponding to tridiagonal functions. We will always assume that the eigenvalues are ordered: $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

To be interesting, doable, and typical, locating parameter values where the eigenvalues coalesce requires these values to be isolated, and this is the reason why we restrict to two parameters in the real symmetric case, and three parameters in the Hermitian case. In fact, as it is well understood (a result dating back to von Neumann and Wigner, [17]), for symmetric functions eigenvalues coalescing is a co-dimension 2 phenomenon; similarly, in the Hermitian case, it is a real co-dimension 3 phenomenon (e.g., see [5] or [15]). Although these results were stated for full, not banded, functions, the stated co-dimensions do not

1991 *Mathematics Subject Classification.* 15A18, 15A23, 65F15, 65F99, 65P30.

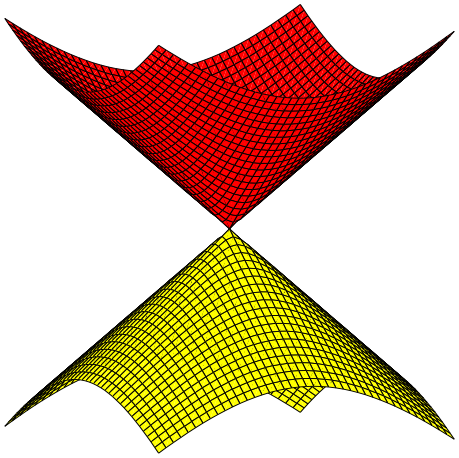
*Key words and phrases.* Coalescing eigenvalues, parameter dependent matrices, tridiagonal, banded, random matrices.

The work was supported in part under INDAM-GNCS, and funds from the Internationalization Plan of the Dep.t of Industrial Engineering of the University of Florence.

change for banded functions with bandwidth $b \geq 1$. This last statement is simple to verify, in the same manner as in [5, Theorem 4.2], upon realizing that the statements $A = QDQ^T$ (symmetric case, $Q$ orthogonal and $D$ diagonal) and $A = UDU^*$ (Hermitian case, $U$ unitary and $D$ diagonal) now entail additional constraints inherited by the structure of $A$; namely, one must have $(QDQ^T)_{ij} = 0$ (respectively, $(UDU^*)_{ij} = 0$), for $i - j > b$ ($j = 1, \ldots, n - b - 1$). Using these extra constraints in the computation of the degrees of freedom one has in specifying $Q$ and $D$ (in the symmetric case), respectively $U$ and $D$ (in the Hermitian case), with a coalescing pair of eigenvalues, and comparing to the degrees of freedom one has in specifying $A$, as in [5], it is immediate to obtain co-dimension 2 in the symmetric banded case, and 3 in the banded Hermitian case.

The co-dimension tells us how many parameters we need in order to expect the occurrence of coalescing; e.g. -generically, in the space of smooth symmetric functions- one needs two parameters in order to observe eigenvalues coalescing.

To witness, for two parameter symmetric functions, one should expect to have isolated parameters' values where the eigenvalues coalesce, eigenvalues surfaces come together at a coalescing point as two upside-down cones (hence, the name *conical intersections*, CIs for short, given to the phenomenon), and coalescing is a robust phenomenon (it persists under small perturbation, though of course the parameters' values where it occurs will typically change). See the figure on the right.



At the same time, we should not expect to have coalescing eigenvalues for functions of one parameter, a fact that precludes being able to locate coalescing eigenvalues by freezing all but one parameter; indeed, for full (i.e., not banded) matrices, we developed numerical methods of topological nature both for symmetric two-parameter functions and Hermitian three-parameter ones. In the symmetric case (see [9]), our technique exploited the relation between coalescing eigenvalues and rotation of eigenvectors along a closed loop containing the coalescing point; the latter property is well established, first observed in [14] and then rediscovered many times. In the Hermitian case, we exploited the relation between coalescing eigenvalues and the preservation, or lack thereof, of a certain eigen-phase as we cover a closed surface enclosing the coalescing point; see [8] for details of our numerical method, and [22] and [2] for fundamental (theoretical) contributions.

The crux of our numerical techniques consists in being able to compute a smooth eigendecomposition along 1-d paths, when the eigenvalues are distinct along the path; this computation we did (and do) with adaptive time-stepping methods, in a way easier appreciated by looking at Figure 1.

At a high level (technical details are explained in [9, 8]), our techniques for full matrices, in the symmetric, respectively Hermitian, cases proceed as follow. We will assume that

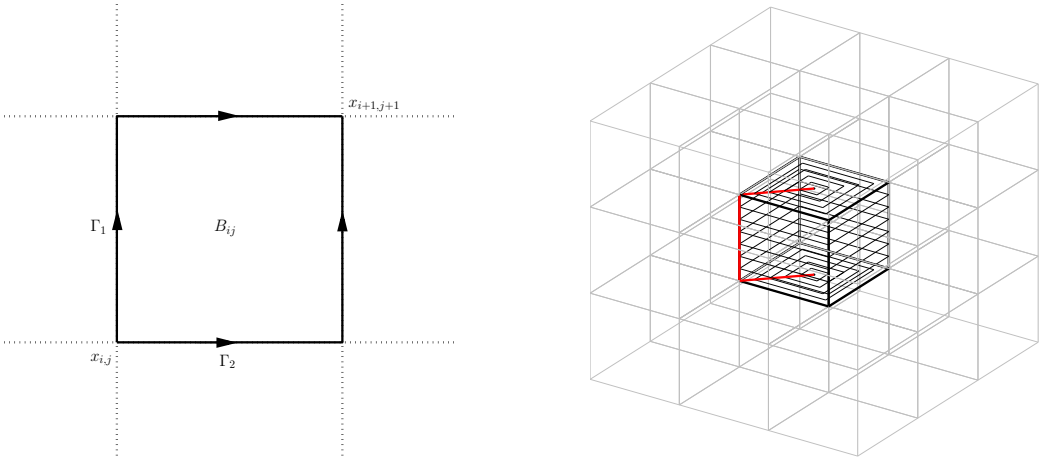



FIGURE 1. Typical domains: 2-d case on left, and 3-d on the right.

the parameter region $\Omega$ of interest is a square in the symmetric case, and a cube in the Hermitian case.

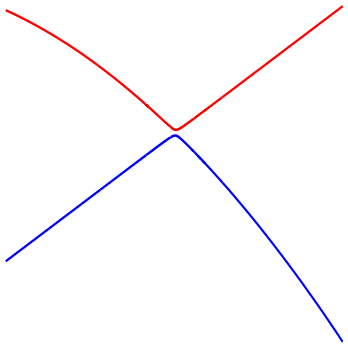
(a) Subdivide $\Omega$ in a number of smaller squares (the subdivision can be done adaptively). On each of these smaller squares, compute smooth eigendecompositions from the SW to the NE corners ($x_{i,j}$ and $x_{i+1,j+1}$ in Figure 1) along the two paths $\Gamma_1$ and $\Gamma_2$, call $U_1$ and $U_2$ the obtained orthogonal factors. Form the diagonal matrix $D = U_1^T U_2$. This matrix $D$ will have 1 or $-1$ on the diagonal, with the $-1$'s betraying if/which eigenvalues have coalesced inside the small square. E.g., suppose we have $n \geq 4$, and $D = \begin{bmatrix} -1 & & & & \\ & 1 & & & \\ & & 1 & & \\ & & & -1 & \\ & & & & \ddots \end{bmatrix}$; then, we can anticipate that inside the square the pairs $(\lambda_1, \lambda_2), (\lambda_2, \lambda_3)$, and $(\lambda_3, \lambda_4)$, have coalesced.

(b) In the Hermitian case, we subdivide $\Omega$ in smaller cubes (again, this can be done adaptively). For each of these small cubes, we proceed from the South to the North poles (the endpoints of the red curve in Figure 1), integrating along the "parallels" (the concentric squares) starting and ending at a point on the red curve; call $s$ the parametrization of the red curve, with $s$ going from 0 to 1. As we move along the meridians, we monitor (smoothly!) the accrued geometric phases $\alpha_j(s)$ (alias Berry phases) for each eigenvalue; to achieve this we must compute the so-called *minimum variation decomposition* (MVD) along the parallels (see [10, 8]). Upon reaching the North pole, and looking at $\alpha_j(1)$, $j = 1, \ldots, n$, we can determine if/which eigenvalues coalesced inside the current cube. Again, see [10, 8] for theoretical and numerical details.

**Remark 1.1.** *The work-horse of our technique is the adaptive integration along the 1-d paths. Referring to* [9, 8] *for specific details, here we simply stress that both eigenvalues and eigenvectors variations impact the choice of adaptive stepsize.*

However, when we directly used the same implementation of [9, 8] for banded functions, we encountered remarkable difficulties and often failed to smoothly continue the 1-d path of eigendecompositions. What happened is that our codes progressively reduced the stepsize, without being able to separate numerically the computed eigenvalues, or to decide that they had coalesced: eventually the stepsize became so small that the codes halted and we were unable to count degeneracies in the region of interest.

The difficulty is not that there are coalescing points along the 1-d path (a fact that we know should not occur), but rather that there is a prevalence of *veering phenomena* (also called *avoided crossings*): some pairs of eigenvalues get very close to one another, without actually coalescing, before eventually separating; yet, the eigenvalues are so close that we cannot distinguish them. See the figure on the right, and see Section 1.2 below for further insight into this fact in the banded case.



For the above reason, in this work we propose and implement a different technique, whereby we dynamically group together close eigenvalues, compute (smoothly) the associated invariant subspace until the veering region is passed, and further rotate this subspace to decide how to continue the eigendecomposition after we exit the veering region. This way of proceeding allowed us to continue the eigendecompositions more reliably than without grouping close eigenvalues, and well beyond the range of applicability of the latter approach. Details are in Section 2.

Finally, in this work, we also present ad-hoc techniques for monitoring eigenvalues coalescing of tridiagonal functions, see Section 2.3.

1.1. **Model problem: degeneracies of banded random matrices depending on parameters.** Random matrices are widely accepted as an important model to obtain statistical properties on the spectra of physical systems; in particular, random matrix models have been successfully applied to obtain statistical information on the spectra of quantum systems. Of particular interest in the random matrix community is the statistical distribution of eigenvalues in the case of full and banded symmetric/Hermitian matrices. In both of these cases, studies exist on the distribution of eigenvalues, and of course of the maximal eigenvalue. The well known semi-circle Wigner's law represents the most striking example of universal property (see [23] for an overview); for the banded case, see [3, 11, 13, 16, 21] for a sample of works, the concern has chiefly been on decay properties of eigenvectors, i.e., "localization/delocalization" properties (see also earlier works by Demko and others, [4], and the recent review [1] for related questions).

An important consideration, for our purposes, is that numerical studies of spectral properties of random matrices are rather simple to do. High level, reliable, software to generate random matrices whose entries obey a particular probability distribution, and to compute their eigenvalues, is easily accessible (e.g., in `Matlab`); and, from these computations, it is possible to extract statistics on the distribution of eigenvalues and the

maximal/minimal eigenvalues (recall that the eigenvalues are continuous functions). This way of proceeding is very well explained in the review article of Edelman and Rao [12]. In other words, through standard numerical experiments, it is easy to gather insight, or to confirm theoretical results, insofar as the eigenvalues distribution for random matrices obeying a certain probability distribution. In the banded case, it is equally standard to numerically compute eigenvalues, though in this case it is less clear which one should be the probability distribution of the entries, see below.

In the full case, studies exist also on distribution of degeneracies, viz. coalescing eigenvalues, for parameter dependent matrices; for example, see the collection of works [24, 25, 26], where remarkable power laws for GOE and GUE (Gaussian Orthogonal and Gaussian Unitary Ensembles, respectively) were derived (our own computational work [8] confirmed these theoretical works). However, in the banded case, no study on statistical properties of degeneracies exists for random parameter dependent matrices, and we venture to say that this is at least in part due to the incredible numerical difficulties of approximating coalescing eigenvalues in this case, and the consequent lack of insight into what one should expect. This task is the one we address in this work.

Generation of random matrix parameter dependent ensembles is itself a delicate task in the banded case. In the full case, generating GOE or GUE ensembles is simple and well documented. For example, following [25], a two parameter GOE ensemble is obtained as

$$A(x,y) = \cos(x)A_1 + \sin(x)A_2 + \cos(y)A_3 + \sin(y)A_4 \ , \tag{1}$$

where $A_1, \ldots, A_4$, are symmetric matrices with diagonal entries in $N(0,1)$ and off-diagonal in $N(0, \frac{1}{2})$; similarly, a three parameter GUE ensemble is obtained as

$$A(x,y,z) = \cos(x)A_1 + \sin(x)A_2 + \cos(y)A_3 + \sin(y)A_4 + \cos(z)A_5 + \sin(z)A_6 \ , \tag{2}$$

where $A_1, \ldots, A_6$, are Hermitian matrices, themselves generated as $A_j = B_j + iC_j$, $j = 1, \ldots, 6$, with $B_j$'s symmetric and $C_j$'s antisymmetric, so that in the end $A_j$'s diagonal entries are in $N(0,1)$ and each off-diagonal entry of the $B_j$'s and $C_j$'s (respecting the relevant symmetries) is in $N(0, \frac{1}{2})$.

However, there is no obvious counterpart on how to generate appropriate ensembles for banded random matrices; e.g., cfr. the differences in [16] and [20]. For example, if one considers a full random symmetric matrix (GOE), as above, and performs on it the standard reduction to tridiagonal form, the end result is a matrix whose diagonal entries are still in $N(0,1)$, but the co-diagonal entries now obey the $\chi$-distribution with degrees of freedom from $n-1$ to 1 (see [12]). Naturally, in a parameter independent setting, one will obtain the same statistical information on the distribution of the eigenvalues whether starting with a full matrix, or with a tridiagonal one, as long as the respective distributions on the entries are adopted. However, in the parameter dependent setting, this way of proceeding is not meaningful; for one thing, it cannot be expected that four (random symmetric) matrices can be simultaneously brought to tridiagonal form; moreover, at best we would learn something about the full, not banded, case (particularly, insofar as the number of degeneracies).

For the above reasons, in our numerical study we consider banded structures obtained by truncating an ensemble generated according to either (1) or (2) (this approach, to the best of our knowledge, was first adopted and studied by Schenker, [20]). We will still refer

to these banded ensembles as GOE and GUE, since orthogonal/unitary transformations preserve the joint element density. Indeed, in this banded case with bandwidth $b$, for the joint element density of a matrix $A$, one has

$$\frac{1}{2^{n/2}}\frac{1}{\pi^{[2n+b(2n-b-1)]/4}}e^{-(\|A\|_F^2)/2}\ ,\ \ b=1,\dots,n-1\ . \tag{3}$$

At the same time, in the banded case, an orthogonal/unitary transformation of a banded symmetric/Hermitian matrix in general destroys the band structure.

For these banded GOE (symmetric case) and GUE (Hermitian case), we are able to give evidence of a power law on the number of degeneracies in terms of the size $n$ of the problem, and of the bandwidth $b$. The power law we find, of the type

$$\#\text{ coalescing } \ =\ cn^p\ , \tag{4}$$

will show several remarkable facts, the most important being the dependence of the exponent $p$ on the bandwidth (also the constant $c$ depends on $b$, and $n$, of course, but this is not our main concern here). For example, in the Hermitian GUE case, the exponent will go from $p = 2.5$ in the full case ($b = n-1$), to $p \approx 3$ in the tridiagonal case ($b = 1$). [For diagonal problems, coalescing occurs along curves for symmetric two-parameter functions, and along 2-d surfaces for Hermitian three-parameter functions.] For values of $b$ in between these two extremal values, we observe a monotonically increasing behavior in $p$ as $b$ decreases[1]. Another striking fact we observe is that the difference in the power law exponent between the GOE and GUE ensembles in the full case, where $p = 2$ for the GOE and $p = 2.5$ for the GUE, progressively disappears as the band approaches $b = 1$, both cases eventually settling toward $p = 3$. Finally, our numerical study in [8] indicated that for full matrices the degeneracies were spatially uniformly distributed, but in the present band case this is surely not the case; this is obvious (see later) in the case $b = 1$ (tridiagonal problems), but also for other band values ($b = 2, 3, \dots$), the spatial distribution of degeneracies does not appear to be uniform, and we conjecture that the spatial distribution converges to being uniform as $b \to n-1$.

1.2. **Numerical issues. Veering and "min-gap".** In the numerical analysis community, it is amply acknowledged that it is hard (if not downright impossible) to decide in finite precision if two eigenvalues are merely close or equal. One of the best known examples of this difficulty is the famous Wilkinson matrix: a symmetric tridiagonal matrix with 1's on the co-diagonal, and diagonal entries given by $(10, 9, \dots, 1, 0, 1, \dots, 9, 10)$. This famous example is easy to write down, but it is otherwise not particularly special as far as having tridiagonal matrices for which one is unable to decide, in finite precision, if eigenvalues are distinct or not; see [27] and [18]. In this work, all computations have been performed in double precision, the default precision in `Matlab`. This corresponds to a machine precision `eps` $\approx 2.2 \times 10^{-16}$.

In the special case of the Wilkinson matrix, since the problem is tridiagonal and unreduced (i.e., no co-diagonal entry is 0), of course Sturm theorem tells us that the eigenvalues must be all distinct.

[1]To be fair, our numerical study is restricted to selected values of $b$, $b = 1, 2, 3, 4, 5$, for values of $n$ up to 150.

**Theorem 1.2** (Sturm)**.** *Given $A = A^T \in \mathbb{R}^{n\times n}$ or $A = A^* \in \mathbb{C}^{n\times n}$, tridiagonal. Then, a necessary condition for two eigenvalues to be equal is that a co-diagonal entry be* 0*.*

We will exploit Theorem 1.2 to find the number of degeneracies in the tridiagonal case; see Section 2.3.

**Remark 1.3.** *For computing eigenvalues of full symmetric/Hermitian matrices, the standard way of proceeding goes through a first step which brings the problem to symmetric real tridiagonal form. In particular, even complex Hermitian unreduced tridiagonal matrices can be taken to real symmetric unreduced tridiagonal form. For this reason, it is tempting to say that –for tridiagonal problems– one may as well restrict to the real symmetric case. However, this is not a correct way to proceed when the interest is in computing coalescing eigenvalues of parameters dependent Hermitian (tridiagonal) functions, a process which requires monitoring the Berry phase along loops. Indeed, transforming a tridiagonal, unreduced, Hermitian function of one parameter (this is what we have along our 1-d paths), into an unreduced, real symmetric function, produces a phase accumulation along the loop which is different from the Berry phase.*

But, beside the tridiagonal case, for which we have developed ad-hoc techniques, our computational techniques for banded matrices with bandwidth $1 < b \ll n-1$, also require special care. This is because of the previously mentioned veering phenomenon, which we now try to elucidate.

As we previously remarked, we must be able to compute a continuous eigendecomposition along 1-d loops. For small bandwidth, there are veering phenomena with consecutive eigenvalues within machine precision, a fact that precludes numerical continuation of the eigendecomposition. During this veering, the associated orthonormal eigenvectors undergo very rapid change (essentially, a clockwise or counter-clockwise rotation by $\pi/2$) within an interval of width less than two consecutive machine numbers. As we show below, this phenomenon gets worse as the dimension $n$ grows, though it can be observed already in $(2, 2)$ systems.

**Example 1.4** (Veering model problem)**.** *This simple* $(2, 2)$ *model clarifies eigenvalues' veering. Take the 2-parameter function*

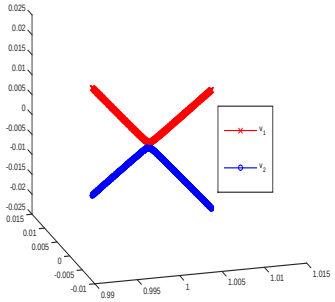
$$A(t) \;=\; \begin{bmatrix} t & ct-1 \\ ct-1 & 2-t \end{bmatrix} \;;\; \textit{with, e.g., } t \in [0,2]\,.$$

*Note, that for $c = 1$ then $A(1) = I$, and if $c \neq 1$, the eigenvalues are distinct for all $t$. If $c \approx 1$, veering occurs at (and near) $t = 1$.*

*Note that the eigenvalues are $\lambda_{1,2} = 1 \pm \left[2 - 2t(c+1) + t^2(c^2+1)\right]^{1/2}$, and the unnormalized orthogonal eigenvectors are*

$$v_1 = \begin{bmatrix} ct-1 \\ \lambda_1 - t \end{bmatrix} \;, \quad v_2 = \begin{bmatrix} ct-1 \\ \lambda_2 - t \end{bmatrix} \;.$$

*Considering $c = 1 \pm \delta$, and $\delta > 0$ small, the eigenvectors seem to exchange with one another, though in fact they do not and sharply rotate by $\pi/2$. In the figure on the right, we show the two eigenvectors for $c = 1 + \delta$ and $t \in [1 - 10\delta, 1 + 10\delta]$, with $\delta = 10^{-3}$.*



It remains to substantiate, and explain, the fact that veering phenomena are prevalent for banded functions, the more so the smaller the bandwidth $b$, and the larger the dimension $n$.

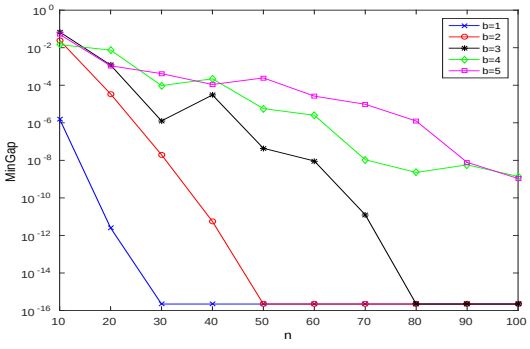
1.2.1. *Min-gap.* Consider the following model problem, first for the real symmetric case, then for the Hermitian case. Note that this model corresponds to the situation we encounter when integrating along the 1-d paths we have[2]. We have a one parameter function $A(t)$, for $t \in [0, 2\pi]$; further, as it is generically true, we can assume that $\lambda_1(t) > \lambda_2(t) > \ldots > \lambda_n(t)$ be the ordered eigenvalues of $A(t)$.

*Symmetric case.* We have a 1-d path of banded matrices, $A(t) = A_1 \cos(t) + A_2 \sin(t) + B \in \mathbb{R}^{n\times n}$ where $t \in [0, 2\pi]$, $B$ is given by $c_2 A_3 + s_2 A_4$, with $c_2^2 + s_2^2 = 1$, but otherwise randomly chosen, and $A_1, A_2, A_3, A_4$, are "banded GOE" matrices with bandwidth $b$. For small $b$, and averaging over several GOE realizations, we monitor the *min-gap* quantity:

$$\texttt{MinGap} := \min_{i=1\ldots,n-1} \min_{t\in[0,\pi]} \lambda_i(t) - \lambda_{i+1}(t) \ . \tag{5}$$

[We stress that the eigendcomposition of the function $A(t)$, $t \in [0, \pi]$, is done with our adaptive solver, whereby the stepsize is adapted to the variation of the eigenvalues and eigenvectors: this is mandatory in order to compute the `MinGap`; in other words, for our purposes, it would be insufficient –and misleading– to sample $A$ at values of $t$ on a fixed grid.]

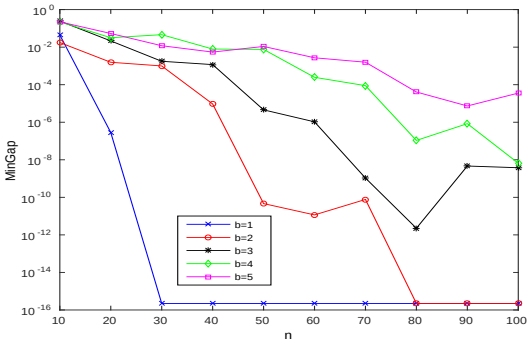
What we observe is that, for $b \ll n$, the `MinGap` decreases as $n$ grows. The speed at which the `MinGap` decreases is itself decreasing as $b$ grows. See the figure on the right (results obtained by averaging over 20 realizations).




[2]The computations on which we report below are themselves quite demanding, particularly those needed for Figure 2.

*Hermitian case.* Now we have a 1-d path of band matrices, $A(t) = A_1\cos(t) + A_2\sin(t) + B \in \mathbb{C}^{n\times n}$ where $t \in [0, 2\pi]$, $B$ is given by $c_2A_3 + s_2A_4 + c_3A_5 + s_3A_6$, with $c_2^2 + s_2^2 = 1$ and $c_3^2 + s_3^2 = 1$, but otherwise randomly chosen, and $A_1, \ldots, A_6$, are "banded GUE" matrices with bandwidth $b$.

Again, we observe that the `MinGap` decreases as $n$ grows, with decreasing speed as $b$ grows. See the figure on the right, for $b = 1, \ldots, 5$, and $n = 10, 20, \ldots, 100$ (again, results obtained by averaging over 20 realizations).




Note that the `MinGap` for banded GUE problems appears to decrease at a slower rate than for banded GOE problems. This is consistent with our results in Section 3, where we observe that the power law governing the number of CIs goes from $\approx n^{2.5}$ (in the full case) to $\approx n^3$ (in the tridiagonal case) for GUE ensembles, and from $\approx n^2$ (in the full case) to $\approx n^3$ (in the tridiagonal case) for GOE ensembles.

Finally, we show (relatively to the banded GOE ensemble, as above) the graphs of the `MinGap` for fixed $n = 100$ in terms of the bandwidth, as well as the converse situation, for fixed $b = 12$ as $n$ increases, see Figure 2 below (results obtained by averaging over 10 realizations).

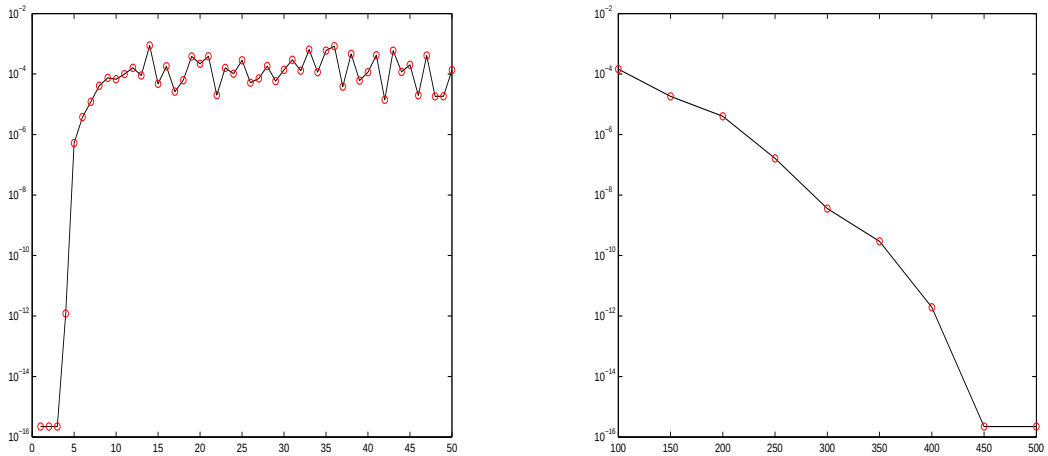


FIGURE 2. `MinGap` banded GOE. Left, $n = 100$, $b = 1, 2, \ldots, 50$. Right, $b = 12$, $n = 100, 150, \ldots, 500$.

To sum up, for given $b \ll n$, and for $n$ sufficiently large, the `MinGap` becomes the size of machine precision. This means that every such problem effectively undergoes a severe veering phenomenon (already for small values of $n$, for small values of $b$), and it is hopeless to try to continue an eigendecomposition like in Algorithm 2.1 below, through a veering

TABLE 1. Average ratio $\rho$ and standard deviation: 1000 conical intersections, $n = 100$.

| $b$ | Average Ratio $\rho$ | Standard Deviation |
|---|---|---|
| 1 | 0.0011 | 0.0194 |
| 2 | 0.0568 | 0.2396 |
| 3 | 0.0705 | 0.1356 |
| 4 | 0.1210 | 0.3484 |
| 5 | 0.1611 | 0.3276 |
| 10 | 0.3487 | 0.2103 |
| 20 | 0.4229 | 0.2081 |
| 50 | 0.4331 | 0.2093 |
| 99 | 0.4296 | 0.2160 |

zone. Specialized tools such as those described in Section 2 are needed to alleviate the impact of veering.

1.2.2. *Ellipses and* `MinGap`. To complete this introduction, we consider the following natural question: What is the mechanism producing small values of `MinGap`? Below, we propose a geometrical explanation, motivated by the work [24].

For simplicity, consider just the symmetric case, though a similar situation holds in the Hermitian case. Let $(\bar{x}, \bar{y})$ be a (isolated) parameter value where two eigenvalues, say $\lambda_k$ and $\lambda_{k+1}$, coalesce, for some $1 \leq k \leq n-1$. If we consider the scalar valued function $f(x, y) := (\lambda_k(x, y) - \lambda_{k+1}(x, y))^2$, then $f(\bar{x}, \bar{y}) = 0$; further, its gradient is also 0 at $(\bar{x}, \bar{y})$. Therefore, locally, the function $f$ is well approximated by a purely quadratic function:

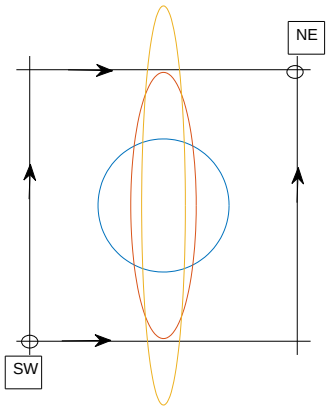
$$f(x, y) \approx h(x, y) := \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}^T H(\bar{x}, \bar{y}) \begin{bmatrix} x - \bar{x} \\ y - \bar{y} \end{bmatrix}$$

for $(x, y) \in B_r(\bar{x}, \bar{y})$, a ball of radius $r$ centered at $(\bar{x}, \bar{y})$; here, $H$ is the Hessian of $f$ (which is positive definite at $(\bar{x}, \bar{y})$, since the conical intersection is isolated). Obviously, the level sets of the function $h$ are ellipses. The shape of these ellipses, that is the ratio of the ellipses semi-axes, is the key aspect to consider. In other words, if we call $\mu_1$ and $\mu_2$ the eigenvalues of the Hessian, with $\mu_1 \geq \mu_2 > 0$, the key quantity to look at is $\rho := \mu_2/\mu_1$. Clearly, $0 < \rho \leq 1$, and the closer $\rho$ is to 1 the more circular is the ellipse, whereas the smaller is $\rho$ the more elongated is the ellipse. Remarkably, the value $\rho$ unambiguously appears to depend on the bandwidth $b$, with smaller values of $\rho$ corresponding to smaller bandwidth $b$ (in the limiting case of $b = 0$, i.e., a diagonal problem, $\rho = 0$ since the Hessian is singular, the conical intersection is not isolated). See Table 1 where we show results of numerical computations averaging the results relative to 1000 conical intersections (randomly selected among the several thousands computed), for banded GOE ensembles, ranging from tridiagonal to full matrices. Clearly, the smaller the bandwidth, the more elongated is the ellipse.

Finally: how does the shape of these ellipses impact prevalence of veering phenomena for banded problems with small bandwidth $b$? An explanation for this fact is based on a simple geometrical argument. Consider a (small) loop in parameter space along which we

compute a smooth eigendecomposition; this is the small square in the earlier description of our topological method. Suppose that inside this small loop there is a coalescing point. Consider the level set of $h$: $h = \lambda_1 u^2 + \lambda_2 v^2$, with $u^2 + v^2 = \delta > 0$; in practice, we can think of $\delta$ as the machine precision `eps`. This level set is an ellipse, more and more elongated the smaller is $b$.

As a consequence, the level sets of $h$ are bound to intersect the loop in parameter space along which we are eigendecomposing the function, the more likely so the smaller is the loop and the value of $b$. See the figure on the right for an exemplification, where we are showing concentric ellipses with $\rho = 1, \frac{1}{4}, \frac{1}{9}$, respectively.




Therefore, even though along the loop the eigenvalues remain distinct, it becomes numerically impossible to distinguish them for small $b$, and a veering phenomenon ensues.

## 2. Dealing with numerical difficulties: Numerical methods

Motivated by the serious numerical difficulties caused by veering phenomena, here we explain the strategy we have adopted to alleviate its impact. We consider a 1-parameter function $A(t)$, for $t \in [0, 1]$, and we are assuming that $\lambda_1(t) > \lambda_2(t) > \ldots > \lambda_n(t)$ are the ordered eigenvalues of $A(t)$. We want to obtain the smooth minimum variation decomposition (MVD), $A(t) = U(t)\Lambda(t)U^*(t)$ with $t \in [0, 1]$ and the eigenvalues ordered along the diagonal of $\Lambda$ (with $Q$ and $Q^T$ instead of $U$ and $U^*$ in the real case).

First, we recall the continuation procedure implemented in the case of well separated eigenvalues. What is meant by this is embodied in the following definition, where notation from Algorithm 2.1 is adopted.

**Definition 2.1.** *Given a full ordered decomposition at $t_j$, a pair of eigenvalues $(\lambda_i, \lambda_{i+1})$ is declared "close to veering" in some interval $[t_j, t_{j+1}]$ if one of the following conditions holds:*

$$\frac{|\lambda_{i+1}(t_{j+1}) - \lambda_i(t_{j+1})|}{|\lambda_i(t_{j+1})| + 1} < \texttt{mindist}\,, \quad \frac{|\lambda_{i+1}^{pred} - \lambda_i^{pred}|}{|\lambda_i^{pred}| + 1} < \texttt{mindist},$$

*where* `mindist` *is a predifined tolerance*[3]*; otherwise, the eigenvalues are said to be "well separated". We do not expect and therefore do not consider the nongeneric case of three or more close eigenvalues.*

[3]e.g., we have used $\texttt{mindist} = 10^6\texttt{eps} \approx 10^{-10}$

In the case of well separated eigenvalues, the continuation procedure is based on a predictor-corrector strategy where the mesh $0 = t_0 < t_1 < \ldots < t_N = 1$ is found adaptively, according to the variation of eigenvalues and eigenvectors. In particular, a relative variation `TolStep` is allowed between predicted and computed factors, for both eigenvalues and eigenvectors. At each step, an ordered eigendecomposition is first computed using standard linear algebra software. Then, since the eigenvectors are uniquely defined up to a phase factor $e^{i\phi}$, with $\phi \in \mathbb{R}$, smoothness is approximately recovered by enforcing minimum variation with respect to the predicted eigenvectors. Below, we denote by $U_j$ the computed approximation to $U(t_j)$. Observe that in the real case $e^{i\phi}$ reduces to $\pm 1$, so that only the signs of the eigenvectors must be corrected and we can recover the exact orthogonal factor $Q(t_j)$. We use predicted factors of the form $U^{pred} = U_j + h\,\dot{U}_j$ and $\Lambda^{pred} = \Lambda_j + h\,\dot{\Lambda}_j$, with approximate derivatives $\dot{U}_j \simeq \dot{U}(t_j)$ and $\dot{\Lambda}_j \simeq \dot{\Lambda}(t_j)$ obtained by replacing $\dot{A}(t_j)$ with $(A(t_{j+1}) - A(t_j))/(t_{j+1} - t_j)$ in the differential equations:

$$\dot{\Lambda} = \mathrm{diag}(U^* \dot{A} U), \quad \dot{U} = UH,$$

where $H$ is the skew-Hermitian matrix such that

$$H_{ik} = -\bar{H}_{ki} = \frac{[U^* \dot{A} U]_{ik}}{\lambda_i - \lambda_k}, \ \text{ for } i < k, \qquad H_{ii} = 0 \ \text{ for } i = 1, \ldots, n.$$

**Algorithm 2.1: Predictor-Corrector step**

Given an ordered eigendecomposition at $t_j$: $A(t_j) = U_j \Lambda(t_j) U_j^*$ and a stepsize $h$, we want the (approximate) MVD at $t_{j+1} = t_j + h$: $A(t_{j+1}) = U_{j+1} \Lambda(t_{j+1}) U_{j+1}^*$.

1. Set $t_{j+1} = t_j + h$, and compute $U^{pred} = U_j + h\,\dot{U}_j$ and $\Lambda^{pred} = \Lambda_j + h\,\dot{\Lambda}_j$;
2. Compute an algebraic ordered Schur decomposition $A(t_{j+1}) = U_c \Lambda(t_{j+1}) U_c^*$;
3. Find the phase matrix $\Phi$ s.t. $\Phi = \mathrm{argmin} ||U^{pred} - U_c \Phi||_F$, and set $U_{j+1} = U_c \Phi$;
4. Set $\rho = \max\{\rho_\lambda, \rho_U\}/$`TolStep` and update $h = h/\rho$, where:
$$\rho_\lambda = \max_i \frac{|\lambda_i(t_{j+1}) - \lambda_i^{pred}|}{|\lambda_i(t_{j+1})| + 1}, \qquad \rho_U = \frac{\|U(t_{j+1}) - U^{pred}\|_F}{\sqrt{n}}$$
5. If $\rho \leq 1.5$, accept the step;
   otherwise declare *failure*, go to step 1, and retry with the new (smaller) $h$.

The above steplength strategy is robust and efficient, as long as the eigenvalues stay well separated, as we observed to be the case for full matrices. But being $\dot{U}$ inversely proportional to the distance between eigenvalues, it can yield prohibitively small values for $h$ in proximity of veering zones, which are routinely encountered in the case of band matrices. In practice, within a veering zone, it is impossible to distinguish a pair of close eigenvalues; to overcome this critical situation, we implemented a strategy (see [6]) in which close eigenvalues are grouped into $2 \times 2$ blocks, and a smooth block-diagonal eigendecomposition is computed until all eigenvalues are well separated again. A key

concern, then, will be how to recover the correct eigendecomposition (smoothly) after we exit the veering zone.

We will distinguish between the case of a real symmetric function $A$ (depending on two real parameters in the original setup), and the case of a complex Hermitian function (depending on three real parameters in the original setup). The following outline of our technique applies to both cases.

(i) Suppose that we have been able to compute successfully the full eigendecomposition of $A$, with ordered eigenvalues, up to a point $t_j$, that $h$ is the current continuation stepsize, and that in the interval $[t_j, t_j + h]$ two consecutive eigenvalues are close to veering, all other eigenvalues being well separated from each other and from this pair. Without loss of generality, assume that these close eigenvalues are the first two eigenvalues: $\lambda_1, \lambda_2$.

(ii) Starting at $t_j$, we compute a block diagonal decomposition with a smooth orthogonal (respectively, unitary) transformation $Q_B$ (respectively, $U_B$):

$$Q_B^T A Q_B = \begin{bmatrix} B & 0 & \dots & 0 \\ 0 & \lambda_3 & \ddots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \lambda_n \end{bmatrix} , \quad t \geq t_j \tag{6}$$

(in the complex case, replace $Q_B^T$ with $U_B^*$). The $(2 \times 2)$ block $B$ has the two eigenvalues $\lambda_1, \lambda_2$, undergoing veering. The computation of $Q_B$ ($U_B$) uses the "smooth continuation of invariant subspaces" technique developed in [7], which is itself based on Riccati transformations.

(iii) We continue with this block eigendecomposition, and monitor the eigenvalues of $B$, until the difference of the two eigenvalues in $B$ has become larger than `mindist`; suppose that this happens at some value $t_f$. Then, we set $t_{j+1} = t_f$, and the issue has become how to compute the full eigendecomposition of $A$, that is how to obtain the eigendecomposition of the block $B$, at $t_{j+1}$; this is done somewhat differently in the real or complex case (see below).

We make a couple of observations.
(a) Note that we are concerned with the case of just two eigenvalues getting close on a small time interval; nevertheless, we also allow the possibility that other pairs of eigenvalues get to within `mindist` while the pair in $B$ is, and in this case the block eigendecomposition is appropriately modified by isolating also these other blocks.
(b) If the two eigenvalues in $B$ remain close for all times past $t_j$, then the block eigendecomposition is carried until the end point.

2.1. **Real symmetric case.** We have $B \in \mathcal{C}^k([t_j, t_{j+1}], \mathbb{R}^{2\times 2})$, symmetric:

$$B = \begin{bmatrix} a & b \\ b & c \end{bmatrix} ,$$

with distinct eigenvalues $\lambda_1 > \lambda_2$ given by:

$$\lambda_{1,2} = \frac{1}{2}\big[a + c \pm \sqrt{\Delta}\big] \ , \ \Delta = (a-c)^2 + 4b^2 \, .$$

We further have that at $t_j$, $B$ has diagonal form: $B(t_j) = \begin{bmatrix} a(t_j)=\lambda_1(t_j) & b(t_j)=0 \\ b(t_j)=0 & c(t_j)=\lambda_2(t_j) \end{bmatrix}$.
To determine the eigendecomposition of $B(t_{j+1})$, we reason as follows.

(i) We know that there is a smooth orthogonal function of eigenvectors $V$ for $B$ on $[t_j, t_{j+1}]$, starting at the identity at $t_j$, and hence it must be a rotation:

$$V = \begin{bmatrix} \cos(\theta(t)) & -\sin(\theta(t)) \\ \sin(\theta(t)) & \cos(\theta(t)) \end{bmatrix} . \tag{7}$$

Also, in a standard way, we have

$$\dot{V} = VH \ , \ H = \begin{bmatrix} 0 & -\dot{\theta} \\ \dot{\theta} & 0 \end{bmatrix} \ , \tag{8}$$

and with a little algebra, we obtain the differential equation satisfied by $\theta$:

$$\dot{\theta} \ = \ \frac{1}{\sqrt{\Delta}}\gamma \ , \ \text{where} \ \ \gamma = \dot{b}\cos(2\theta) - \frac{\dot{a}-\dot{c}}{2}\sin(2\theta) \, . \tag{9}$$

(ii) At $t_{j+1}$, we know that the orthogonal matrix $Z$:

$$Z = \frac{1}{\sqrt{(\lambda_1 - c)^2 + b^2}} \begin{bmatrix} \lambda_1 - c & -b \\ b & \lambda_1 - c \end{bmatrix} \tag{10}$$

is such that $\lambda_1 - c \geq 0$, and

$$Z^T B(t_{j+1}) Z = \begin{bmatrix} \lambda_1(t_{j+1}) & 0 \\ 0 & \lambda_2(t_{j+1}) \end{bmatrix} \ .$$

Moreover, being $V$ a rotator, we must always have

$$V(t_{j+1}) = Z \quad \text{or} \quad V(t_{j+1}) = -Z \ . \tag{11}$$

**Algorithm**. Our algorithm determines the way that eigenvectors rotate, and accordingly fixes the sign of the approximation for $V(t_{j+1})$ in (11), by looking at the sign of $\dot{\theta}$. From (9), for $\theta(t_j) = 0$, we have $\dot{\theta}(t_j) = \frac{1}{\sqrt{\Delta(t_j)}}\dot{b}(t_j)$, so the sign of $\dot{\theta}(t_j)$ is that of $\dot{b}(t_j)$, and the rotation will be clockwise if $\dot{b}(t_j) < 0$ and counterclockwise if $\dot{b}(t_j) > 0$ (the case of $\dot{b}(t_j) = 0$ would require us to look at higher derivatives, but it is of no present concern). Now, since $b(t_j) = 0$, to decide whether $\dot{b}(t_j) < 0$ or $\dot{b}(t_j) > 0$, in practice, we look at $b$ for a few consecutive values in between $[t_j, t_{j+1}]$, to determine if $\theta$ is increasing or decreasing, and then use the form of $Z$ in (10) as the appropriate form of $V(t_{j+1})$. [Note that our choice boils down to having $\theta \in (-\pi/2, \pi/2)$. This is bound to be correct, especially since we begin grouping the eigenvalues near the veering point].

The overall algorithm allows us to tackle problems with close eigenvalues, well beyond the range of applicability of the direct eigendecomposition approach: loosely speaking, the direct eigendecomposition approach of Algorithm 2.1 fails as soon as eigenvalues cannot be distinguished in finite precision (see the discussion on the `MinGap` in section 1.2.1), whereas the present approach, based on grouping and ungrouping of close eigenvalues,

and on fixing the sign in (11) based upon the behavior of $b$, works for as long as the function $b$ is meaningfully distinguishable from 0, and hence its sign is unambiguously detected.

2.2. **Complex Hermitian case.** In the Hermitian case, things are more complicated, though even in this case we have a mean to alleviate the impact of veering, and the overall algorithm is somewhat similar to that in the real symmetric case.

We have $B \in \mathcal{C}^k([t_j, t_{j+1}], \mathbb{C}^{2\times 2})$, smooth, Hermitian:

$$B = \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix} , \quad b = b_r + ib_i \ , \ a, c, b_r, b_i \in \mathbb{R} \, ,$$

with distinct eigenvalues for all $t \in [t_j, t_{j+1}]$, $\lambda_1 > \lambda_2$, given by:

$$\lambda_{1,2} = \frac{1}{2}\big[a + c \pm \sqrt{\Delta}\big] \ , \ \Delta = (a-c)^2 + 4\,|b|^2 \ .$$

We further have $B(t_j) = \begin{bmatrix} a = \lambda_1 & 0 \\ 0 & c = \lambda_2 \end{bmatrix}_{t=t_j}$, and will assume that the function $b(t) \neq 0$ for $t > t_j$, in particular that $\dot{b}(t_j) \neq 0$.

We need to characterize the unitary function $U$ of eigenvectors of $B$, giving the MVD of $B$, where $U = I$ at $t_j$. Standard arguments give:

$$\dot{U} \ = \ UH \ , \ H = \begin{bmatrix} 0 & H_{12} \\ -\bar{H}_{12} & 0 \end{bmatrix} \ , \ H_{12} = \frac{(U^*\dot{B}U)_{12}}{\lambda_2 - \lambda_1} \, ,$$

which (since $\lambda_2 - \lambda_1 = -\sqrt{\Delta}$) can be rewritten as

$$\dot{U} \ = \ U \begin{bmatrix} 0 & -\frac{1}{\sqrt{\Delta}}(U^*\dot{B}U)_{12} \\ \frac{1}{\sqrt{\Delta}}(U^*\dot{B}U)_{21} & 0 \end{bmatrix} . \tag{12}$$

Next, note that

$$\frac{d}{dt} \det U = (\det U)\ \mathrm{trace}\big(U^* \frac{dU}{dt}\big) = 0$$

and therefore $\det U$ is constant, and thus it is equal to 1. This means that $U$ must have the form

$$U \ = \ \begin{bmatrix} \alpha & -\bar{\beta} \\ \beta & \bar{\alpha} \end{bmatrix} \ , \ |\alpha|^2 + |\beta|^2 = 1 \, .$$

Therefore, we will look for $U$ of the form

$$U \ = \ \begin{bmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} =: D_1 V \, , \tag{13}$$

and now we reason as follows.

(a) From (13), we observe that $V$ is real valued; therefore, from the relation $U^*BU = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} =: \Lambda$, and the form of $U$, we must have

$$V^T\big(D_1^{-1}BD_1\big)V = \Lambda \ , \qquad \text{or} \qquad V^TCV = \Lambda \ , \quad \text{with} \quad C := D_1^{-1}BD_1 \ , \tag{14}$$

which implies that the function $C$ must be real valued, and symmetric. This means that the role of the phase $\phi$ is to make the complex valued function $b$ real. Indeed, note that if $C = C^T$ and real valued, then forming $C$ explicitly gives that we must have

$$C = \begin{bmatrix} e^{-i\phi/2} & 0 \\ 0 & e^{i\phi/2} \end{bmatrix} \begin{bmatrix} a & b \\ \bar{b} & c \end{bmatrix} \begin{bmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{bmatrix} = \begin{bmatrix} a & \delta \\ \delta & c \end{bmatrix} ,$$

with

$$\delta = e^{-i\phi} b = e^{i\phi} \bar{b} \in \mathbb{R} . \tag{15}$$

(b) For the real symmetric problem relative to $C$, the two eigenvalues $\lambda_1$ and $\lambda_2$ undergo veering, just like in the real symmetric case of Section 2.1, and the computational technique proceeds similarly to the case there. Therefore

$$V = \frac{1}{\sqrt{(\lambda_1 - c)^2 + \delta^2}} \begin{bmatrix} \lambda_1 - c & -\delta \\ \delta & \lambda_1 - c \end{bmatrix},$$

and from (13)

$$U = \frac{1}{\sqrt{(\lambda_1 - c)^2 + \delta^2}} \begin{bmatrix} (\lambda_1 - c)\, e^{i\phi/2} & -\delta\, e^{i\phi/2} \\ \delta\, e^{-i\phi/2} & (\lambda_1 - c)\, e^{-i\phi/2} \end{bmatrix}.$$

Recalling (15) we then have

$$U = \frac{1}{\sqrt{(\lambda_1 - c)^2 + |b|^2}} \begin{bmatrix} \lambda_1 - c & -b \\ \bar{b} & \lambda_1 - c \end{bmatrix} \begin{bmatrix} e^{i\phi/2} & 0 \\ 0 & e^{-i\phi/2} \end{bmatrix}.$$

(c) Finally, we need to specify the way that the angle $\phi$ brings $b$ on the real axis (i.e., clockwise or counterclockwise). To do this, observe that $b(t) = b(t_j) + (t - t_j)\dot{b}(t_j) + \ldots$, and thus (since $b(t_j) = 0$) we look at the vector $(\dot{b}_r(t_j), \dot{b}_i(t_j))$ and rotate $b(t)$ onto the positive/negative real axis depending on whether $\dot{b}_r(t_j)$ is positive or negative. In practice, we look at a few consecutive values of $b$ to assess $\phi$.

With this, the algorithm in the complex case is fully described. Again, by using the grouping/ungrouping strategy described, we have been able to alleviate the impact of veerings, and to solve problems which were unaccessible to a direct eigendecomposition approach.

2.3. **Tridiagonal case.** The case of a tridiagonal function is special, and allows us to devise taylor-made techniques in order to locate CIs. We have

$$A = \begin{bmatrix} a_1 & b_2 & & & \\ b_2 & a_2 & b_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b_n \\ & & & b_n & a_n \end{bmatrix} \in \mathbb{R}^{n\times n} , \text{ or } A = \begin{bmatrix} a_1 & b_2 & & & \\ \bar{b}_2 & a_2 & b_3 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & b_n \\ & & & \bar{b}_n & a_n \end{bmatrix} \in \mathbb{C}^{n\times n}, \tag{16}$$

where the entries in the real case depend on two (real) parameters, and in the complex case depend on three (real) parameters.

The basic idea is to exploit Theorem 1.2: "eigenvalues coalescing can only occur if a co-diagonal entry is 0." Now, in the real case, this means that, for some $i = 2, \ldots, n$, we must have $b_i(x, y) = 0$; generically, this defines a curve in the $(x, y)$-parameter space. In the complex case, for some $i = 2, \ldots, n$, we must have $\mathrm{Re}(b_i(x, y, z)) = 0$ and $\mathrm{Im}(b_i(x, y, z)) = 0$; again, generically, these two equations define a curve in the $(x, y, z)$-parameter space.

Looking at the function $A$, we will have $n - 1$ such curves. Along each of these curves, the problem becomes reduced: $A = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$ (of course, the dimensions of $A_1$ and $A_2$ depend on the particular index $i$ for which $b_i = 0$). We can thus monitor the eigenvalues of $A_1$ and $A_2$ (as we move along a specific curve) to locate parameter values where they coalesce. The latter task we perform by bisection.

In short, our task has become:

(1) Compute the curves $b_i = 0$, $i = 2, \ldots, n$.
(2) For each of the above curves, locate CIs (using bisection) by monitoring the eigenvalues of the decoupled blocks.

Both of the above tasks can be carried out by somewhat classical numerical methods, of course taking care of the periodicity we have for our model tridiagonal GOE and GUE ensembles in (1) and (2).

**Remark 2.2.** *It would surely be convenient to have the curves $b_i = 0$ as "functions," so that the bisection procedure can refine the search for CIs as desired. Now, this wish can be fulfilled in the 2-parameter symmetric GOE case,* (1)*, simply because the co-diagonal function which we are zeroing can be written as $R_1 \cos(x - \alpha) + R_2 \cos(y - \beta) = 0$, and we can solve for one of $x$ or $y$ in terms of the other. However, in the complex GUE case,* (2)*, it is not possible to write the curves analytically; for this reason, we compute these curves by pseudo-arclength continuation, after having located an initial point on the curve, and we use very small continuation steps, so that the bisection procedure will later be able to accurately locate the CIs.*

## 3. Numerical computations: banded GOE and GUE ensembles

Our implementations follow closely the descriptions in the previous sections, with the basic approach being to continue a complete eigendecomposition for as long as feasible, and use a block eigendecomposition when the eigenvalues are so close that computing a complete eigendecomposition becomes not possible.

We stress that, although we can and do assume the eigenvalues to be distinct along the paths we follow, in practice while we continue a complete eigendecomposition they are often so close that we witness repeated failures during integration, simply because we reach the minimum value of the stepsize. Indeed, the goal of grouping near eigenvalues is in order to reduce the number of failures during integration. In fact, in our experience, a failure leads to a miscount of CIs, and repeated failures lead to a severe miscount of the number of CIs. Thus, we need to achieve two goals: avoid reaching the minimum allowed stepsize, and being able to ungroup the eigenvalues before having completed the loop. These goals are controlled by two input values: `hmin` and `mindist`. We note that when we do not group nearby eigenvalues, quite often we reach `hmin`, whereas if the value used for `mindist` is too large, then grouped eigenvalues remain paired for the entire loop.

In all of our experiments, we used `hmin` $= 10^{-14}$ or $10^{-15}$. Through several experiments, we calibrated the value of `mindist` and eventually chose it to be equal to $10^6$`eps`, since this value resulted in the smallest number of failures, hence in the most reliable value for the obtained number of CIs.

Also, we used `TolStep` $= 10^{-2}$ as the error tolerance during continuation along the paths (see Algorithm 2.1). Finally, by looking at the form of the functions in (1) and (2), quite clearly we can exploit the periodicity with respect to one of the parameters. For this reason, for the GOE (1) we used a grid of $512 \times 1024$ square boxes on the domain $[0, \pi] \times [0, 2\pi]$ (thus, we always integrated along segments of length $\approx 0.0061$), whereas for the GUE (2) we used a grid of $2n \times 2n \times n$ boxes for a matrix function in $\mathbb{C}^{n\times n}$ on the domain $[0, 2\pi] \times [0, 2\pi] \times [0, \pi]$. For each value of $n$, we report on results obtained by averaging over 10 different realizations.

All computations have been performed on the FoRCE Research Computing Environment cluster available at Georgia Tech, currently equipped with 1008 six core CPUs.

**Remark 3.1.** *In spite of the improvements we achieved by grouping, in the GOE case with bandwidth 2 we were not able to successfully complete all the experiments for dimensions larger than 70. We believe this is due to the fact that, in those cases, a significant number of events (rotations) we are trying to capture occur entirely within regions whose size is smaller then machine precision. Such behavior is consistent with the value observed for the* `MinGap` *in Section 1.2.1.*

The results in the first two tables, Table 2 and Table 3, are about the performance of the algorithms. The most noteworthy feature in these tables is that the grouping of eigenvalues is prevalent for small bandwidth and it becomes less needed as soon as $b = 5$ (for the given values of $n$). Another interesting observation is that the percentage of steps rejected is much the same across the values of $n$ and $b$. Finally, please observe the different values of $n$ for the case of $b = 2$ in the GOE case.

In Figure 3 we show the average number of CIs in the tridiagonal cases, for both GOE and GUE. The approximated value of $p$ in the power law expressing the number of CIs, see (4), is observed to be the same (at 4 digits accuracy) for GOE and GUE cases.

Figures 4 and 5 show the number of CIs in function of the dimension and bandwidth, for the GOE and GUE cases, respectively, with the corresponding values of $p$ in the power law; again, see (4). The case of $b = 2$ for the GOE is singled out, because of the different values of $n$ we used. Note that these figures report on the number of degeneracies for the domains we used (which are half of the periodic square/cube where (1-2) are defined).

Finally, Table 4 reports the computed values of $p$ and of the constant $c$ in the power law (4), $cn^p$, for several values of $b$. The values reported for $b = n - 1$ are exact (in the limit of large $n$, of course), from [24] for the GOE and [26] for the GUE.

## 4. Conclusions

In this work we have considered banded symmetric, respectively Hermitian, matrix functions depending on 2, respectively 3, real parameters. Our main concern has been that of detecting parameters' values where the eigenvalues became equal (these are also called degeneracies, or conical intersections).

TABLE 2. The table shows (left to right): bandwidth, dimension of the problem, average and relative standard deviation of the number of conical intersections detected, total number of steps taken (in millions), percentage of steps rejected, number of times that eigenvalues have been grouped per million of steps. All values are an average over 10 realizations from the "banded GOE" ensemble. We stress that these numbers refer to the domain $[0, \pi] \times [0, 2\pi]$.

| b | dim | avg CI | rsd | steps | rej | group |
|---|---|---|---|---|---|---|
| 2 | 30 | 1547 | 6.8 | 29.0 | 1.4 | $1.9 \times 10^2$ |
| | 40 | 3628 | 4.7 | 32.6 | 3.4 | $3.6 \times 10^3$ |
| | 50 | 6840 | 4.5 | 39.9 | 5.8 | $3.1 \times 10^4$ |
| | 60 | 11443 | 3.4 | 48.9 | 7.4 | $9.1 \times 10^4$ |
| | 70 | 17329 | 3.4 | 59.7 | 8.2 | $2.1 \times 10^5$ |
| 3 | 50 | 4307 | 6.0 | 30.0 | 2.1 | $2.4 \times 10^2$ |
| | 60 | 6893 | 3.9 | 33.0 | 3.8 | $2.8 \times 10^3$ |
| | 70 | 10510 | 2.3 | 37.2 | 5.4 | $1.1 \times 10^4$ |
| | 80 | 15240 | 3.2 | 44.0 | 7.3 | $3.0 \times 10^4$ |
| | 90 | 21332 | 4.2 | 52.4 | 8.7 | $6.8 \times 10^4$ |
| | 100 | 27605 | 2.8 | 60.6 | 9.9 | $1.1 \times 10^5$ |
| | 110 | 37089 | 5.0 | 74.4 | 11.3 | $2.2 \times 10^5$ |
| | 120 | 46354 | 2.4 | 86.8 | 12.0 | $3.7 \times 10^5$ |
| 4 | 50 | 3237 | 3.3 | 27.9 | 0.6 | 1.6 |
| | 60 | 5267 | 4.3 | 28.7 | 1.2 | 22 |
| | 70 | 7962 | 1.6 | 30.4 | 2.4 | $4.4 \times 10^2$ |
| | 80 | 11241 | 3.9 | 32.6 | 3.6 | $2.3 \times 10^3$ |
| | 90 | 15385 | 2.7 | 35.7 | 5.1 | $4.9 \times 10^3$ |
| | 100 | 20620 | 3.3 | 40.2 | 6.9 | $1.1 \times 10^4$ |
| | 110 | 26260 | 3.2 | 45.5 | 8.7 | $2.1 \times 10^4$ |
| | 120 | 33554 | 3.9 | 52.9 | 10.3 | $5.8 \times 10^4$ |
| 5 | 50 | 2890 | 5.6 | 27.5 | 0.2 | 0.2 |
| | 60 | 4446 | 3.3 | 27.7 | 0.4 | 1.4 |
| | 70 | 6666 | 3.2 | 28.2 | 0.8 | 3.2 |
| | 80 | 9234 | 3.3 | 29.0 | 1.5 | 58 |
| | 90 | 12600 | 3.5 | 30.1 | 2.3 | $3.2 \times 10^2$ |
| | 100 | 16715 | 4.4 | 32.1 | 3.6 | $8.6 \times 10^2$ |
| | 110 | 20955 | 3.4 | 34.7 | 5.1 | $3.0 \times 10^3$ |
| | 120 | 26649 | 3.2 | 37.4 | 6.6 | $7.5 \times 10^3$ |

As motivation of our numerical study, and application of it, we have been studying banded GOE and GUE ensembles, and obtained evidence of a power law on the number of degeneracies in terms of the size $n$ of the problem, and of the bandwidth $b$: $cn^p$, where $p$ was found to be a function of $b$, varying from $p = 2$ to $p \approx 3$ for GOE ensembles as the matrix went from being full to being tridiagonal, and going from $p = 2.5$ to $p \approx 3$ in the case of GUE ensembles.

TABLE 3. The table shows (left to right): bandwidth, dimension of the problem, average and relative standard deviation of the number of conical intersections detected, total number of steps taken (in millions) along meridians and parallels, percentage of steps rejected along meridians and parallels, number of times eigenvalues have been grouped. All values are an average over 10 realizations from the "banded GUE" ensemble. We emphasize that these numbers refer to the domain $[0, 2\pi] \times [0, 2\pi] \times [0, \pi]$.

| b | dim | avg CI | rsd | steps mer | steps par | rej mer | rej par | group |
|---|---|---|---|---|---|---|---|---|
| 2 | 50 | 14807 | 4.0 | 14.4 | 1197.2 | 0.7 | 9.0 | $1.3 \times 10^5$ |
| | 55 | 19664 | 4.3 | 19.2 | 1830.5 | 0.7 | 9.8 | $4.4 \times 10^5$ |
| | 60 | 25087 | 3.1 | 24.9 | 2591.2 | 0.7 | 10.3 | $9.4 \times 10^5$ |
| | 65 | 31638 | 3.4 | 31.6 | 3698.9 | 0.7 | 10.6 | $2.4 \times 10^6$ |
| | 70 | 38683 | 2.7 | 39.6 | 5120.8 | 0.9 | 10.9 | $5.1 \times 10^6$ |
| | 75 | 46516 | 3.7 | 48.8 | 6677.0 | 1.0 | 11.0 | $8.7 \times 10^6$ |
| | 80 | 57294 | 4.0 | 59.1 | 9098.4 | 1.4 | 11.1 | $1.5 \times 10^7$ |
| 3 | 50 | 12511 | 4.4 | 14.2 | 697.3 | 0.4 | 3.5 | $1.4 \times 10^1$ |
| | 55 | 16947 | 3.6 | 19.0 | 989.1 | 0.5 | 4.5 | $6.7 \times 10^2$ |
| | 60 | 20884 | 2.0 | 24.6 | 1344.6 | 0.5 | 5.1 | $2.8 \times 10^3$ |
| | 65 | 26493 | 3.5 | 31.3 | 1819.4 | 0.5 | 5.8 | $2.0 \times 10^4$ |
| | 70 | 32632 | 2.5 | 39.1 | 2389.3 | 0.5 | 6.5 | $5.6 \times 10^4$ |
| | 75 | 38644 | 2.5 | 48.1 | 3074.4 | 0.5 | 6.9 | $2.0 \times 10^5$ |
| | 80 | 46555 | 2.5 | 58.4 | 3977.9 | 0.5 | 7.6 | $5.2 \times 10^5$ |
| 4 | 50 | 11665 | 3.5 | 14.2 | 599.2 | 0.3 | 0.9 | 0.8 |
| | 55 | 15107 | 2.9 | 18.9 | 804.1 | 0.3 | 1.3 | 1.6 |
| | 60 | 19107 | 2.3 | 24.5 | 1071.3 | 0.3 | 1.6 | 7.6 |
| | 65 | 23732 | 2.7 | 31.2 | 1393.5 | 0.3 | 1.8 | 7.8 |
| | 70 | 29147 | 2.7 | 39.0 | 1771.3 | 0.4 | 2.4 | $1.9 \times 10^2$ |
| | 75 | 34888 | 2.1 | 47.9 | 2230.7 | 0.3 | 2.8 | $1.1 \times 10^3$ |
| | 80 | 41972 | 2.0 | 58.2 | 2754.6 | 0.4 | 3.2 | $5.1 \times 10^2$ |
| 5 | 50 | 11428 | 1.6 | 14.2 | 580.0 | 0.3 | 0.3 | 0 |
| | 55 | 14644 | 2.9 | 18.9 | 762.9 | 0.3 | 0.4 | 0 |
| | 60 | 17983 | 4.4 | 24.5 | 1009.5 | 0.3 | 0.5 | 1.4 |
| | 65 | 22668 | 2.4 | 31.1 | 1310.5 | 0.3 | 0.7 | 0 |
| | 70 | 27447 | 2.8 | 38.9 | 1616.7 | 0.3 | 0.7 | 2 |
| | 75 | 32981 | 2.1 | 47.8 | 2006.9 | 0.3 | 0.9 | 0.9 |
| | 80 | 39389 | 2.5 | 58.0 | 2432.1 | 0.3 | 1.1 | 23 |

The outstanding task in computing these conical intersections was to overcome the difficulties caused by close (but not identical) eigenvalues, a prevalent phenomenon for small bandwidth. We elucidated this difficulty (aka as veering of eigenvalues) and then developed algorithms that adaptively group together close eigenvalues until these veering zones are bypassed. So doing, we managed to distinguish the occurrences of degeneracies well beyond the limitations imposed by having close eigenvalues, before eventually succumbing to the inherent constraints of finite precision.

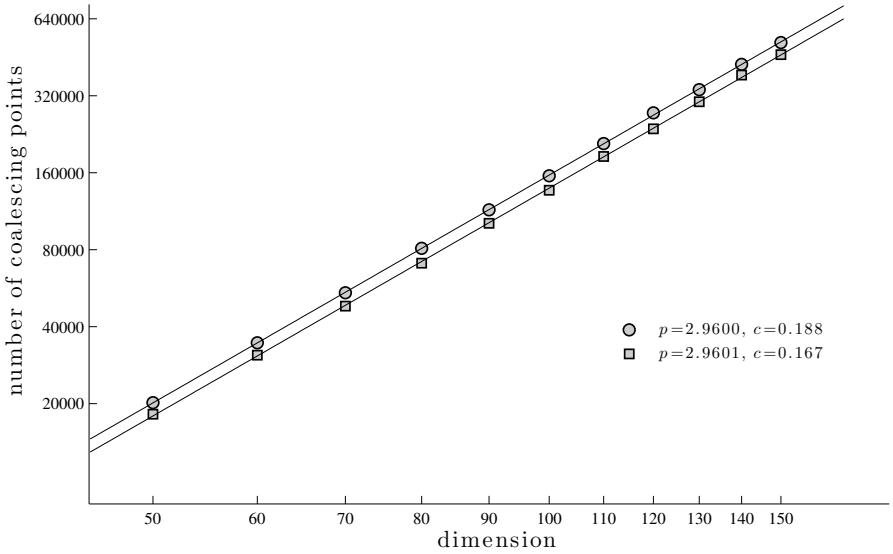



FIGURE 3. Average number of CIs computed in the tridiagonal GOE (circles) and GUE (square) cases. Figure also shows the slope of the linear regression in log-log scale.

TABLE 4. The table collects the power laws, see (4).

| | bandwidth | power law $p$ | coefficient $c$ |
|---|---|---|---|
| GOE | 1 | 2.96 | 0.188 |
| | 2 | 2.85 | 0.191 |
| | 3 | 2.73 | 0.194 |
| | 4 | 2.66 | 0.193 |
| | 5 | 2.55 | 0.269 |
| | $\vdots$ | $\vdots$ | $\vdots$ |
| | $n-1$ | 2 | $\pi/2$ |
| GUE | 1 | 2.96 | 0.167 |
| | 2 | 2.85 | 0.437 |
| | 3 | 2.77 | 0.508 |
| | 4 | 2.72 | 0.563 |
| | 5 | 2.63 | 0.760 |
| | $\vdots$ | $\vdots$ | $\vdots$ |
| | $n-1$ | 2.5 | $512/(135\sqrt{3\pi})$ |

## REFERENCES

[1] M. BENZI, P. BOITO AND N. RAZOUK, Decay properties of spectral projectors with applications to electronic structure. *SIAM Review*, 55-1, pp. 3-64, 2013.

[2] M. V. BERRY, Quantal phase factors accompanying adiabatic changes. *Proc. Roy. Soc. Lond.*, A392, pp. 45–57, 1984.

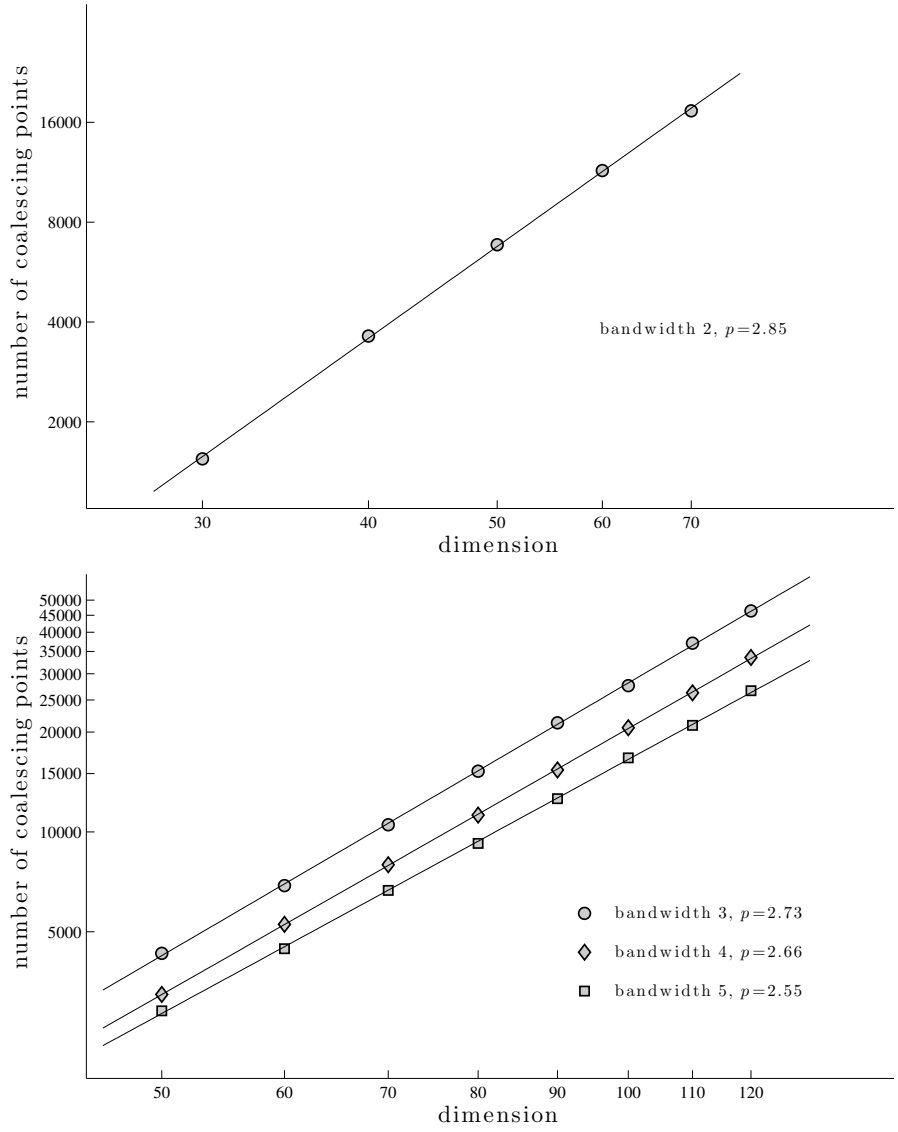



FIGURE 4. Average number of CIs computed in the "banded GOE" case. Top: bandwidth 2; Bottom: bandwidth 3 (circles), 4 (diamonds), 5 (squares). Figures also show the slopes of the linear regressions for each bandwidth in log-log scale.

[3] G. CASATI, B.V. CHIRIKOV, I. GUARNERI AND F.M. IZRAILEV, Band-random-matrix model for quantum localization in conservative systems. *Phys. Rev. E*, 48-3, pp. 1613-1616, 1993.

[4] S. DEMKO, W. MOSS AND P. SMITH, Decay rates for inverses of band matrices. *Math. Comp.*, 43-168, pp. 491-499, 1984.

[5] L. DIECI AND T. EIROLA, On smooth orthonormal factorizations of matrices. *SIAM J. Matrix Anal. Appl.*, 20, pp. 800-819, 1999.

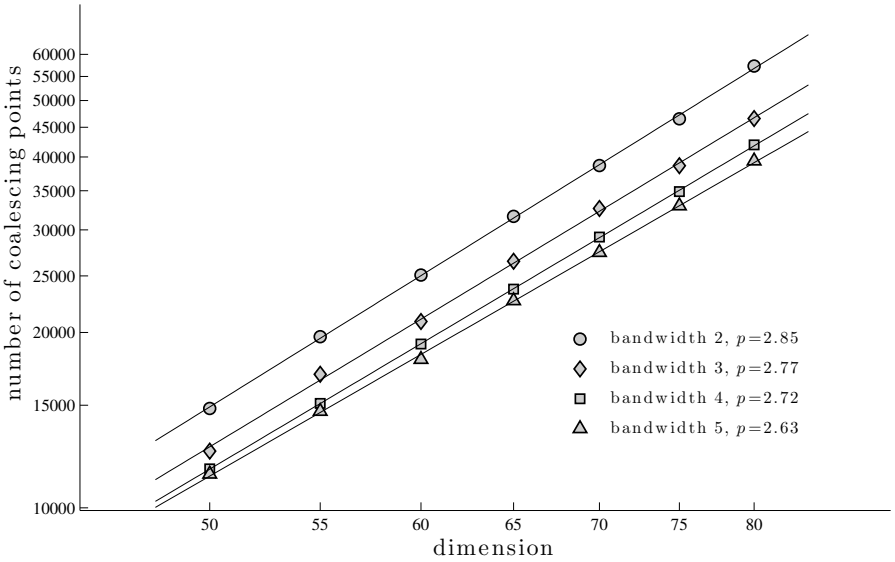



FIGURE 5. Average number of CIs computed in the "banded GUE" case. Bandwidth 2 (circles), 3 (diamonds), 4 (squares), 5 (triangles). Figures also show the slopes of the linear regressions for each bandwidth in log-log scale.


[6] L. DIECI, M.G. GASPARO AND A. PAPINI, Path following by SVD. *Lecture Notes in Computer Science*, vol. 3994, pp. 677-684, Springer, New York, 2006.

[7] L. DIECI AND A. PAPINI, Continuation of eigendecompositions. *Future Generation Computer Systems*, 19, pp. 1125-1137, 2003.

[8] L. DIECI, A. PAPINI AND A. PUGLIESE, Approximating coalescing points for eigenvalues of Hermitian matrices of three parameters. *SIAM J. Matrix Anal. Appl.*, 34-2, pp. 519-541, 2013.

[9] L. DIECI AND A. PUGLIESE, Singular values of two-parameter matrices: An algorithm to accurately find their intersections. *Mathematics and Computers in Simulation*, 79-4, pp. 1255–1269, 2008.

[10] L. DIECI AND A. PUGLIESE, Hermitian matrices depending on three parameters: Coalescing eigenvalues. *Linear Algebra and its Applications*, 436, pp. 4120–4142, 2012.

[11] M. DISERTORI, H. PINSON, T. SPENCER, Density of states for random band matrix. *Comm. Math. Phys.*, 232, pp.83, 2002.

[12] A. EDELMAN, N.R. RAO, Random matrix theory. *Acta Numerica*, pp. 233-297, 2005.

[13] Y.V. FYODOROV AND A.D. MIRLIN, Scaling properties of localization in random band matrices: a $\sigma$-model approach. *Phy. Rev. Letters*, 67-18, pp. 2405-2409, 1991.

[14] G. HERNZBERG AND H.C. LONGUET-HIGGINS, Intersection of potential energy surfaces in polyatomic molecules. *Disc. Faraday Soc.*, 35:77–82, 1963.

[15] J. KELLER, Multiple eigenvalues. *Linear Algebra and its Applications*, 429, pp. 2209-2220, 2008.

[16] M. KUS, M. LEWENSTEIN AND F. HAAKE, Density of eigenvalues of random band matrices. *Phys. Rev. A*, 44-5, pp. 2800-2808, 1991.

[17] J. VON NEUMANN AND E. WIGNER, Eigenwerte bei adiabatischen prozessen. *Physik Zeitschrift*, 30, pp. 467–470, 1929.

[18] B. PARLETT AND C. VÖMEL, The spectrum of a glued matrix. *SIAM J. Matrix Anal. Appl.*, 31-1, pp. 114–132, 2009.

[19] L. PASTUR AND M. SHCHERBINA, *Eigenvalue Distribution of Large Random Matrices*. AMS Mathematical Surveys and Monographs, v. 171, Providence RI, 2011.

[20] J. Schenker, Eigenvector localization for random band matrices with power law band width. *Comm. Math. Phys.*, 290, pp.1065–1097, 2009.
[21] S. Sodin, The spectral edge of some random band matrices. *Annals of Mathematics*, 172, pp. 2223-2251, 2010.
[22] A. J. Stone, Spin-Orbit Coupling and the Intersection of Potential Energy Surfaces in Polyatomic Molecules. *Proc. Roy. Soc. Lond.*, A351:141–150, 1976.
[23] C. A. Tracy and H. Widom, The Distributions of Random Matrix Theory and their Applications. *New Trends in Mathematical Physics*, Editor: V. Sidoravicius, pp 753–765, 2009.
[24] M. Wilkinson and E.J. Austin, Densities of degeneracies and near-degeneracies. *Physical Review A*, 47-4, pp. 2601–2609, 1993.
[25] P.N. Walker and M.J. Sanchez and M. Wilkinson, Singularities in the spectra of random matrices. *J. Mathem. Phys.*, 37-10, pp. 5019–5032, 1996.
[26] P.N. Walker and M. Wilkinson, Universal fluctuations of Chern integers. *Physical Review Letters*, 74-20, pp. 4055–4058, 1995.
[27] Q. Ye, On close eigenvalues of tridiagonal matrices. *Numer. Math.*, 70, pp. 507–514, 1995.

School of Mathematics, Georgia Institute of Technology, Atlanta, GA 30332 U.S.A.
*E-mail address*: dieci@math.gatech.edu

Dept. of Industrial Engineering, Univ. of Florence, viale G. Morgagni 40-44, 50134 Florence, Italy
*E-mail address*: alessandra.papini@unifi.it

Dept. of Mathematics, Univ. of Bari "A. Moro," Via Orabona 4, 70125 Italy
*E-mail address*: alessandro.pugliese@uniba.it