Version of April 1^{st}, 2014

HERMITIAN MATRICES OF THREE PARAMETERS: PERTURBING COALESCING EIGENVALUES AND A NUMERICAL METHOD

LUCA DIECI AND ALESSANDRO PUGLIESE

Abstract. In this work we consider Hermitian matrix valued functions of 3 (real) parameters, and are interested in generic coalescing points of eigenvalues, *conical intersections*. Unlike our previous works [7, 4], where we worked directly with the Hermitian problem and monitored variation of the geometric phase to detect conical intersections inside a sphere-like region, here we consider the following construction: (i) Associate to the given problem a real symmetric problem, twice the size, all of whose eigenvalues are now (at least) double, (ii) perturb this enlarged problem so that –generically– each pair of consecutive eigenvalues coalesce along curves, and only there, (iii) analyze the structure of these curves, and show that there is a small curve, nearly planar, enclosing the original conical intersection point. We will rigorously justify all of the above steps. Furthermore, we propose and implement an algorithm following the above approach, and illustrate its performance in locating conical intersections.

Notation. Below, $\Omega \subset \mathbb{R}^3$ indicates an open region of $\mathbb{R}^3$ diffeomorphic to the open unit ball; $\xi = (\alpha, \beta, \gamma) \in \Omega$ will indicate a general point in Ω. The metric is the Euclidean metric. We write $A \in \mathcal{C}^k(\Omega, \mathbb{C}^{n\times n})$, $k \geq 1$, to indicate a smooth complex matrix-valued function defined on Ω and further $A \in \mathcal{C}^\omega$ if the dependence on parameter(s) is analytic. We write A^* for the conjugate transpose of a matrix A, and have $A = A^*$ for a Hermitian matrix. The word Hermitian will imply complex valued entries. Similarly, the word symmetric will be restricted to matrices with real entries. With $\sigma(A)$ we indicate the set of eigenvalues (repeated by multiplicity) of A.

1. Introduction

In this paper we consider a Hermitian matrix depending on three real parameters, α, β, γ:

$$A = B + iC \ , \quad A^* = A \in \mathbb{C}^{n\times n} \ , \ \ B, C \in \mathbb{R}^{n\times n} \ , \qquad \text{so that} \qquad B^T = B \ , \ \ C^T = -C \ ,$$

where $A \in \mathcal{C}^k(\Omega, \mathbb{C}^{n\times n})$, $k \geq 1$. Naturally, the eigenvalues of A can be taken as continuous functions of $\xi \in \Omega$, and ordered as $\mu_1(\xi) \geq \mu_2(\xi) \geq \cdots \geq \mu_n(\xi)$, for all $\xi \in \Omega$, which we can assume to be always the case. It is well known that, in general, the eigenvalues of A

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

Key words and phrases. Hermitian matrix valued functions, conical intersection, perturbed eigenproblem, diabolical points.

This work was supported in part under INDAM GNCS. Support from the School of Mathematics of the Georgia Institute of Technology is also gratefully acknowledged.

do not enjoy any extra smoothness. In fact –generically– they are not differentiable where they coalesce, regardless of their labeling (ordering). This is in sharp contrast to the case of Hermitian functions depending on one real parameter, where an important theorem of Rellich (see [12]) tells us that the eigenvalues can be chosen to remain at least $\mathcal{C}^1$ functions even if they coalesce, though of course one must allow for eigenvalues to exchange their ordering upon coalescing. However, as we said, in the present multiparameter case, when eigenvalues coalesce they are, generically, merely continuous functions (irrespective of their labeling), and come together at the coalescing point in a double cone like fashion, hence the name of *conical intersection* points (CIs, for short) given to parameter values where eigenvalues coalesce [1]. This phenomenon has been studied extensively, and we refer to [7] for background on the theory, references to this topic, and justification of the terminology adopted in the present work. Presently, we simply recall that, for a Hermitian matrix function, having a pair of coalescing eigenvalues is a real codimension 3 phenomenon; that is, generically: one needs three real parameters to observe it, the phenomenon occurs at isolated parameter values, and it persists upon perturbation (of course, occurring at perturbed parameter values). It follows that it is a generic property for a Hermitian function of three real parameters to have eigenvalues that coalesce at isolated points given by CIs. Likewise, for real symmetric matrix functions, having coalescing eigenvalues is a codimension 2 phenomenon (a fact known since [22]), and therefore, generically, a real symmetric function of two real parameters will have coalescing eigenvalues at isolated CI points, and one of three real parameters will have coalescing eigenvalues along curves of CI points in parameter space. We stress that, because of the codimension of the phenomena under interest, one cannot locate CI's by working with one parameter at a time, as curves of eigenvalues of a symmetric/Hermitian function will generically not intersect.

Remark 1.1. Locating CIs is not only an interesting and challenging task, but it is also a problem of great relevance in the physical and engineering sciences. For example, it plays a key role in chemical physics [1, 2, 24], in random matrix theory [23], and in structural dynamics [8, 18], among others.

From the mathematical point of view, given the singularity nature of generic CIs, it is natural to study what happens for a perturbed problem. Of course, perturbing within the class of Hermitian functions will merely change the location (in parameter space) of the CI point. For this reason, much in the spirit of singularity theory, the approach recently discussed in [3, 14, 15, 16] has lead the authors to a generic non-Hermitian perturbation. In this case, a generic CI point gets replaced by a ring of so-called "exceptional points," whereby one no longer has a full basis of eigenvectors. Unfortunately, the non-Hermitian, non-diagonalizable, problem is harder to treat numerically than the symmetric eigenproblem. For this reason, our approach is to study a perturbed problem where sufficient symmetry is retained in a way that is conducive to the development of a robust and stable computational method.

[1] In the Physics literature, these points are often called *diabolical points* or *degeneracies*.

Our interest in this paper is two-fold: (a) To study a symmetric eigenproblem, which results from a perturbation of the symmetric matrix associated to the original Hermitian one, so that the original CIs are now replaced by curves of CI points, and rigorously study the structure of these curves near a CI of the original Hermitian problem; and, based upon this mathematical theory, (b) to propose a novel technique to localize and approximate conical intersection points of the original Hermitian function.

In [4], we developed a numerical method to approximate CIs, based on the theoretical development of [7], and ultimately based on the remarkable geometrical insight of Stone and Berry; see [21, 2]. The approach considered in this paper, instead, is perturbative in nature and is based on the idea below.

1.1. **Basic Idea/Approach.**

(1) Enlarge A to a symmetric real-valued problem twice the size:

$$M = \begin{bmatrix} B & -C \\ C & B \end{bmatrix} \ , \tag{1.1}$$

so that $M = M^T \in \mathcal{C}^k(\Omega, \mathbb{R}^{2n\times 2n})$.

• *Concern.* Note that the function M is a very special type of symmetric function of three parameters, since all eigenvalues of M appear in repeated pairs:

$$\begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} V & W \\ W & -V \end{bmatrix} = \begin{bmatrix} V & W \\ W & -V \end{bmatrix} \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} .$$

In particular, if $\nu_1 = \nu_2 \geq \nu_3 = \nu_4 \geq \cdots \geq \nu_{2n-1} = \nu_{2n}$ indicate the eigenvalues of M, then $\nu_{2k-1} = \nu_{2k} = \mu_k$, $k = 1, \dots, n$, for all $x \in \Omega$, and at the isolated parameter values where two eigenvalues of the original problem coalesced (say, $\mu_1 = \mu_2$), now we must have a quadruplet of coalescing eigenvalues of M (say, $\nu_1 = \nu_2 = \nu_3 = \nu_4$); these types of coalescings are highly nongeneric properties for symmetric functions of three parameters.

(2) Perturb M (see below for how the perturbation matrix is chosen):

$$M \to M + \varepsilon E \ , \ \ E^T = E \, , \tag{1.2}$$

where E is a constant matrix of norm 1 and ε is a small (positive) number. The matrix E is needed so that for the function $M + \varepsilon E$ the eigenvalues coalesce along curves (a generic property for symmetric functions of three parameters). That is, if $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{2n}$ indicate the eigenvalues of $M + \varepsilon E$, now we will generically have

$$\lambda_1 = \lambda_2, \ \lambda_2 = \lambda_3, \ \lambda_3 = \lambda_4, \ \dots \ ,$$

along non-intersecting curves in parameter space.

(3) We want to rigorously show that –locally– breaking of the CI of the original problem, that is of the quadruplet of coalescing eigenvalues of M, gives three curves for the perturbed problem, and that these curves come near one another around the original CI. In Figure 1 (which is relative to the results of a practical computation), we show the following situation:

(i) $\lambda_1 = \lambda_2 \longrightarrow$ "red curve;"
(ii) $\lambda_3 = \lambda_4 \longrightarrow$ "blue curve;"
(iii) $\lambda_2 = \lambda_3 \longrightarrow$ "green curve."

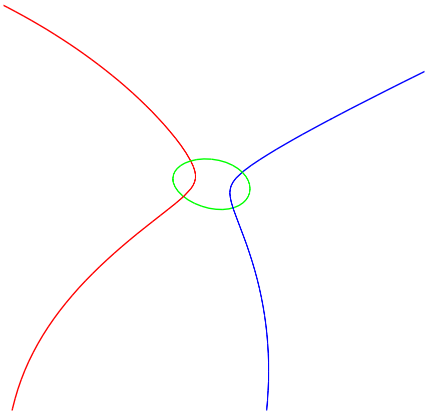

FIGURE 1. The red, blue, and green, curves.

Our main theoretical goal in this paper will be to rigorously show that Figure 1 is qualitatively correct, and the green curve is a (nearly planar) small closed curve encircling the blue and red curves near the original CI.

(4) Based upon the theoretical development above, we will propose a new method to approximate the CI points. In much simplified terms, the method will consist in using a path-following algorithm to follow the red and blue curves while monitoring also points on the green curve, as well as other branches of red/blue curves. Eventually, a refinement procedure will be used to approximate the CI point starting from the point(s) which have been detected on the green curve(s). Details are in Section 4.

A plan of this paper is as follows. In the next Section, we will show that, near an original CI point, the structure of M and of the perturbation E can be taken of a simplified form, so that locally the problem is reduced to a (generic) (4×4) symmetric problem of the type $M + \varepsilon E$, with M and E of an appropriate form (see Theorem 2.10); we note that, in general, now both the (4×4) function M and the "perturbation" E will depend on all of the problem's parameters. [We emphasize right away that this simplified form is only for the purpose of theoretical analysis, in the algorithmic development we will never explicitly build these simplified forms.] In Section 3, we first show how for the (4×4) perturbed symmetric problem $M + \varepsilon E$ we can justify the claim above about the red, blue, and green curves, under the further assumption that the (4×4) perturbation matrix E is constant, and M is independent of ε. Then, we prove that the situation for the general problem is effectively much the same. In Section 4 we propose a new technique to approximate CI points of the original problem by following the approach outlined in Section 1.1, and show its performance on an Example. Finally, in Section 5 we give concluding remarks, and in the Appendix we give complete proofs of the key result, Theorem 2.10, used in Section 3.

2. Simplify Structure

Here we show that, near a CI point, the structure of the function M and of the perturbation E in (1.2) can be chosen of an appropriately simplified form.

In what follows, we assume that $\xi \in \Omega$ is an isolated coalescing point for A. Without loss of generality, we assume $\mu_1(\xi) = \mu_2(\xi)$. Furthermore, since ξ is an isolated coalescing point, we let R to be a parallelepiped, containing ξ, and inside which there are no further coalescings of eigenvalues.

2.1. **Simplify** M.

First, we simplify M, thanks to the block-diagonalization result of Hsieh and Sibuya (see [11], and also [9]).

Theorem 2.1. *Let $A = A^* \in \mathcal{C}^k(\Omega, \mathbb{C}^{n\times n})$, $k \geq 1$, $A = B + iC$, and let $M = \begin{bmatrix} B & -C \\ C & B \end{bmatrix}$. In R, there exists orthogonal $Q \in \mathcal{C}^k(R, \mathbb{R}^{2n\times 2n})$, such that*

$$Q^T M Q = \begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix} \ , \ M_1 \in \mathbb{R}^{4\times 4} \ , \ M_2 \in \mathbb{R}^{2n-4,2n-4}\,, \tag{2.1}$$

and further $\sigma(M_1) \cap \sigma(M_2) = \emptyset$ and the coalescing quadruplet of $\sigma(M)$ (i.e., $\nu_1 = \nu_2 = \nu_3 = \nu_4$ which equals $\mu_1 = \mu_2$) is in $\sigma(M_1)$.

Proof. We use the orthogonal function Z, guaranteed to exist from [11], such that $Z^*AZ = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$, $A_1 \in \mathbb{C}^{2\times 2}$, $A_2 \in \mathbb{C}^{n-2,n-2}$, and $\sigma(A_1) \cap \sigma(A_2) = \emptyset$ in R, with $\sigma(A_1) = \{\mu_1, \mu_2\}$. Further, $A_k^* = A_k$, $k = 1, 2$, and so $A_k = B_k + iC_k$, $B_k^T = B_k$, $C_k^T = -C_k$, $k = 1, 2$. In particular, $B_1 = \begin{bmatrix} a & b \\ b & d \end{bmatrix}$, $C_1 = \begin{bmatrix} 0 & c \\ -c & 0 \end{bmatrix}$. Now, if $Z = U + iV$, then

$$\begin{bmatrix} U & V \\ V & -U \end{bmatrix}^T \begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} U & V \\ V & -U \end{bmatrix} = \begin{bmatrix} B_1 & 0 & C_1 & 0 \\ 0 & B_2 & 0 & C_2 \\ -C_1 & 0 & B_1 & 0 \\ 0 & -C_2 & 0 & B_2 \end{bmatrix}.$$

Now, take the permutation matrix Π, $\Pi = \begin{bmatrix} I_2 & 0 & 0 & 0 \\ 0 & 0 & I_{n-2} & 0 \\ 0 & I_2 & 0 & 0 \\ 0 & 0 & 0 & I_{n-2} \end{bmatrix}$, and observe that

$$\Pi^T \begin{bmatrix} U & V \\ V & -U \end{bmatrix}^T \begin{bmatrix} B & -C \\ C & B \end{bmatrix} \begin{bmatrix} U & V \\ V & -U \end{bmatrix} \Pi = \begin{bmatrix} B_1 & C_1 & 0 & 0 \\ -C_1 & B_1 & 0 & 0 \\ 0 & 0 & B_2 & C_2 \\ 0 & 0 & -C_2 & B_2 \end{bmatrix}.$$

So, the claim is verified with the matrix $Q = \begin{bmatrix} U & V \\ V & -U \end{bmatrix} \Pi$, which is clearly orthogonal, since $Z^*Z = I_n$. □

At this point, without loss of generality, we can assume that the coalescing point for A occurs at the origin, $(\alpha,\beta,\gamma)=(0,0,0)$ and that no other coalescing of eigenvalues of A occur at the origin nor in a neighborhood of it. So, for M_1 we have the structure

$$M_1=\begin{bmatrix} a & b & 0 & c \\ b & d & -c & 0 \\ 0 & -c & a & b \\ c & 0 & b & d \end{bmatrix}.$$

We can further simplify this, by shifting it by $\frac{a+d}{2}$: $M_1 \to M_1-\frac{a+d}{2}I$ and obtain a modified M_1 with zero trace:

$$M_1=\begin{bmatrix} \frac{a-d}{2} & b & 0 & c \\ b & \frac{d-a}{2} & -c & 0 \\ 0 & -c & \frac{a-d}{2} & b \\ c & 0 & b & \frac{d-a}{2} \end{bmatrix}. \tag{2.2}$$

Definition 2.2. We say that $(\alpha,\beta,\gamma)=(0,0,0)$ is a generic CI for M_1 if the system $\begin{cases} a-d=0 \\ b=0 \\ c=0 \end{cases}$ is satisfied at $(\alpha,\beta,\gamma)=(0,0,0)$, and the Jacobian

$$J=\begin{bmatrix} \partial_\alpha((a-d)/2) & \partial_\beta((a-d)/2) & \partial_\gamma((a-d)/2) \\ \partial_\alpha b & \partial_\beta b & \partial_\gamma b \\ \partial_\alpha c & \partial_\beta c & \partial_\gamma c \end{bmatrix}_{(0,0,0)}$$

is invertible.

Remark 2.3. We point out (see [7]) that $(0,0,0)$ is a generic CI for M_1 if and only if it is a generic CI for the Hermitian function

$$\begin{bmatrix} \frac{a-d}{2} & b+ic \\ b-ic & -\frac{a-d}{2} \end{bmatrix}.$$

Henceforth, we will always work under the following assumption:

Assumption 2.4. We assume that $\xi=(0,0,0)$ is an isolated generic CI point for A, with $\mu_1(\xi)=\mu_2(\xi)$.

Finally, since the above Jacobian J is invertible, on account of the inverse function theorem we change variables in M_1, and let:

$$x=\frac{a-d}{2}\ ,\ y=b\ ,\ z=-c\ .$$

Summary 2.5. In conclusion, in a neighborhood of the origin, we can assume to be working with $M=\begin{bmatrix} M_1 & 0 \\ 0 & M_2 \end{bmatrix}$, $\sigma(M_1)\cap\sigma(M_2)=\emptyset$, and the following simplified function

M_1:

$$M_1 = \begin{bmatrix} x & y & 0 & -z \\ y & -x & z & 0 \\ 0 & z & x & y \\ -z & 0 & y & -x \end{bmatrix}, \tag{2.3}$$

where x, y, z are $\mathcal{C}^k$ functions of (α, β, γ) in a neighborhood of the origin, vanish at the origin, and can be used as a local coordinates system.

2.2. **Choosing E.**

Here we discuss the type of perturbation matrix E that we will consider when taking $M + \varepsilon E$. [Step (2) of the outline given in Section 1.1].

We will take the perturbation matrix to be of the type

$$E = \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix} \ , \ E_1^T = E_1 \ , \ E_2^T = E_2 \ , \ E_{1,2} \in \mathbb{R}^{n\times n} \ . \tag{2.4}$$

Remark 2.6. Clearly (2.4) is not the most general form of a symmetric matrix. However, we have the freedom to choose E, and we elect to choose it as in (2.4). Moreover, in our context, E essentially is equivalent to having taken a general perturbation. In fact, suppose we take a general symmetric perturbation matrix $E = E^T = \begin{bmatrix} E_{11} & E_{12} \\ E_{12}^T & E_{22} \end{bmatrix}$ with all blocks being in $\mathbb{R}^{n\times n}$. Then, we can rewrite

$$E = \begin{bmatrix} E_1 + F_1 & E_2 - F_2 \\ E_2 + F_2 & -E_1 + F_1 \end{bmatrix}, \qquad \text{where}$$

$$E_1 = \frac{E_{11} - E_{22}}{2}, \ E_2 = \frac{E_{12} + E_{12}^T}{2}, \ F_1 = \frac{E_{11} + E_{22}}{2}, \ F_2 = \frac{E_{12}^T - E_{12}}{2} \ .$$

Looking at $M + E$, with M as in (1.1), we can rewrite

$$M + E = \begin{bmatrix} B + F_1 & -(C + F_2) \\ C + F_2 & B + F_1 \end{bmatrix} + \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix},$$

and the first of these two matrices on the right-hand-side has the same structure of M, and can thus be absorbed into it.

Remark 2.7. Notice that we cannot interpret E in (2.4) as coming from a perturbation (in $\mathbb{C}^{n\times n}$) of the original A. This observation supports our claim in the Introduction that we are not perturbing the original Hermitian problem. Ultimately, this is the reason why we are able to work with a symmetric function.

Before proceeding, we give a simple result on the spectrum of $M + E$.

Lemma 2.8. *For M as in* (1.1) *and E as in* (2.4), *we always have $\sigma(M+E) = \sigma(M-E)$, and therefore $\sigma(M + E) = -\sigma(-M + E)$.*

Proof. We have

$$(M{+}E)\begin{bmatrix} v \\ w \end{bmatrix} = \lambda \begin{bmatrix} v \\ w \end{bmatrix} \iff \begin{cases} (B+E_1)v-(C-E_2)w=\lambda v \\ (C+E_2)v+(B-E_1)w=\lambda w \end{cases} \iff (M{-}E)\begin{bmatrix} w \\ -v \end{bmatrix} = \lambda \begin{bmatrix} w \\ -v \end{bmatrix} .$$

The final inference follows from this, since $\sigma(M+E) = -\sigma(-M-E) = -\sigma(-M+E)$. □

2.3. **Blocking the perturbed problem.** Next, suppose we have blocked M with a function Q as in Section 2.1, see Theorem 2.1 and (2.1), and let E be a matrix of the structure given in (2.4). Then, we have the following result.

Lemma 2.9. *Let Q be as in* (2.1)*, and $E = \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix}$, $E_j^T = E_j$, $j = 1, 2$. Then,*

$$\begin{aligned} &Q^TEQ = \begin{bmatrix} \widehat{E} & F \\ F^T & H \end{bmatrix},\ \widehat{E} = \begin{bmatrix} \widehat{E}_{11} & \widehat{E}_{12} \\ \widehat{E}_{12} & -\widehat{E}_{11} \end{bmatrix},\ \widehat{E}_{11}^T = \widehat{E}_{11} \in \mathbb{R}^{2\times 2},\ \widehat{E}_{12}^T = \widehat{E}_{12} \in \mathbb{R}^{2\times 2}\ , \\ &H = \begin{bmatrix} H_{11} & H_{12} \\ H_{12} & -H_{11} \end{bmatrix}\ ,\ H_{11}^T = H_{11} \in \mathbb{R}^{n-2,n-2}\ ,\ H_{12}^T = H_{12} \in \mathbb{R}^{n-2,n-2}\ , \\ &F = \begin{bmatrix} F_{11} & F_{12} \\ F_{12} & -F_{11} \end{bmatrix}\ ,\ F_{11}, F_{12} \in \mathbb{R}^{2,n-2}\ . \end{aligned}$$

Proof. Recall (see the proof of Theorem 2.1) that the function Q has the form $Q = \begin{bmatrix} U & V \\ V & -U \end{bmatrix}\Pi$, where $Z = U + iV$ is unitary. Therefore,

$$\begin{bmatrix} U & V \\ V & -U \end{bmatrix}^T \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix} \begin{bmatrix} U & V \\ V & -U \end{bmatrix} =: \begin{bmatrix} \widehat{E}_1 & \widehat{E}_2 \\ \widehat{E}_2 & -\widehat{E}_1 \end{bmatrix}\ ,\ \widehat{E}_1^T = \widehat{E}_1\ ,\ \widehat{E}_2^T = \widehat{E}_2\ .$$

Using the block permutation Π completes the proof. □

So, we really have the following structure:

$$Q^T(M+\varepsilon E)Q\ =\ \begin{bmatrix} M_1+\varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2+\varepsilon H \end{bmatrix}\ , \tag{2.5}$$

where M_1, M_2, $\widehat{E}$, F and H are as in Lemma 2.9, and –by construction– they are $\mathcal{C}^k$ functions in a neighborhood of the origin.

The essential ingredient to obtain the sought result, that is to validate Figure 1, is the following Theorem, where we smoothly block-diagonalize the function in (2.5), while retaining separate spectra for the diagonal blocks.

Theorem 2.10. *Let M_1, M_2, $\widehat{E}$, F and H be as in* (2.5)*, with M_1, M_2 as in Summary 2.5, and $\widehat{E}$, F and H as in Lemma 2.9. Then, there exists $\varepsilon_0 > 0$ sufficiently small, and a neighborhood U_0 of the origin in $\mathbb{R}^3$, such that for $\varepsilon \in J_0 = (-\varepsilon_0, \varepsilon_0)$, and $(\alpha, \beta, \gamma) \in U_0$ the following hold.*

There exist near the identity Q, orthogonal, smooth in $(\alpha,\beta,\gamma)\in U_0$, and analytic for $\varepsilon\in J_0$, such that

$$Q^T\begin{bmatrix} M_1+\varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2+\varepsilon H\end{bmatrix}Q \;=\; \begin{bmatrix}\widetilde{M_1} & 0 \\ 0 & \widetilde{M_2}\end{bmatrix}$$

where

$$\begin{aligned} \widetilde{M_1} &= [M_1+\sum_{k=1}^{\infty}\varepsilon^{2k}E_{2k}]+\varepsilon[\widehat{E}+\sum_{k=1}^{\infty}\varepsilon^{2k}E_{2k+1}]\,, \\ \widetilde{M_2} &= [M_2+\sum_{k=1}^{\infty}\varepsilon^{2k}H_{2k}]+\varepsilon[H+\sum_{k=1}^{\infty}\varepsilon^{2k}H_{2k+1}]\,, \end{aligned} \tag{2.6}$$

and $\sigma(\widetilde{M_1})\cap\sigma(\widetilde{M_2})=\emptyset$.

Moreover, for all $k=1,2,\dots$, the functions E_{2k} all have the same structure as M_1 does, and the functions E_{2k+1} all have the same structure as $\widehat{E}$ itself. Similarly for the functions H_{2k} and H_{2k+1}. All of these functions are $\mathcal{C}^k$ functions of (α,β,γ) in U_0, and are analytic in ε.

More precisely, we can write $\widetilde{M_1}=\widetilde{M}+\varepsilon\widetilde{E}$, where

$$\begin{aligned} \widetilde{M} &= \begin{bmatrix} \widetilde{\alpha} & \widetilde{\beta} & 0 & -\widetilde{\gamma} \\ \widetilde{\beta} & \widetilde{\delta} & \widetilde{\gamma} & 0 \\ 0 & \widetilde{\gamma} & \widetilde{\alpha} & \widetilde{\beta} \\ -\widetilde{\gamma} & 0 & \widetilde{\beta} & \widetilde{\delta}\end{bmatrix}\,, \quad \textit{and} \\ \widetilde{E} &= \begin{bmatrix}\widetilde{E}_1 & \widetilde{E}_2 \\ \widetilde{E}_2 & -\widetilde{E}_1\end{bmatrix}\,, \quad \widetilde{E}_1=\begin{bmatrix}\widetilde{a} & \widetilde{b} \\ \widetilde{b} & \widetilde{c}\end{bmatrix}\,, \widetilde{E}_2=\begin{bmatrix}\widetilde{d} & \widetilde{e} \\ \widetilde{e} & \widetilde{f}\end{bmatrix}\,, \end{aligned} \tag{2.7}$$

where $\widetilde{a},\widetilde{b},\widetilde{c},\widetilde{d},\widetilde{e},\widetilde{f}$, as well as $\widetilde{\alpha},\widetilde{\beta},\widetilde{\gamma},\widetilde{\delta}$, are all $\mathcal{C}^k$ functions of (α,β,γ) in U_0, and are analytic in ε; indeed, all of these functions admit an expansion in powers of ε^2, say $\widetilde{a}=a+\sum_{k=1}^{\infty}a_k\varepsilon^{2k}$, and similarly for the other functions, for $\varepsilon\in J_0$.

Proof. See Appendix. □

3. The green, blue and red curves

Our concern is to study the geometrical structure of the set of points where eigenvalues of $M+\varepsilon E$ coalesce. Relatively to this problem, we are now going to validate Figure 1, that is the structure of the blue, red, and green curves.

3.1. **Simplified M, constant E.** We begin by considering the problem under the following simplifying assumption: "we assume M to be as in (2.3), and E is constant". That is:

$$\begin{aligned} M &= \begin{bmatrix} M_1 & M_2 \\ -M_2 & M_1 \end{bmatrix}, \quad M_1 = \begin{bmatrix} x & y \\ y & -x \end{bmatrix}, M_2 = \begin{bmatrix} 0 & -z \\ z & 0 \end{bmatrix}, \\ E &= \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix}, \quad E_1 = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, E_2 = \begin{bmatrix} d & e \\ e & f \end{bmatrix}, \end{aligned} \tag{3.1}$$

where $a, b, c, d, e, f \in \mathbb{R}$.

So, let M and E be as in (3.1), and let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ be the eigenvalues of $M + \varepsilon E$. The next Lemma gives a result on the eigenvalues/eigenvectors for E of the previous form and it will be useful to further simplify the problem.

Lemma 3.1. *Consider any symmetric matrix $E \in \mathbb{R}^{4\times 4}$ of the form in* (3.1)*:*

$$E = \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix}, \quad E_1 = \begin{bmatrix} a & b \\ b & c \end{bmatrix}, E_2 = \begin{bmatrix} d & e \\ e & f \end{bmatrix}.$$

Then, the eigenvalues of E are of the form $\lambda = \{\pm\kappa_1, \pm\kappa_2\}$, where we can assume $\kappa_1 \geq \kappa_2 \geq 0$. Moreover, if $\begin{bmatrix} v \\ w \end{bmatrix}$ (with $v, w \in \mathbb{R}^2$) is an eigenvector associated to the eigenvalue $\lambda = \kappa_1$ (or $\lambda = \kappa_2$), then $\begin{bmatrix} -w \\ v \end{bmatrix}$ is an eigenvector associated to $-\lambda$.

Proof. Let λ be an eigenvalue of E, and $E \begin{bmatrix} v \\ w \end{bmatrix} = \lambda \begin{bmatrix} v \\ w \end{bmatrix}$. From the form of E, this means

$$\left\{ \begin{array}{l} E_1 v + E_2 w = \lambda v \\ E_2 v - E_1 w = \lambda w \end{array} \right. \iff \begin{bmatrix} E_1 & E_2 \\ E_2 & -E_1 \end{bmatrix} \begin{bmatrix} -w \\ v \end{bmatrix} = -\lambda \begin{bmatrix} -w \\ v \end{bmatrix}.$$

□

The following is a consequence of Lemma 3.1.

Corollary 3.2. *Let $\kappa_1 \geq \kappa_2 \geq -\kappa_2 \geq -\kappa_1$ be the eigenvalues of E in Lemma 3.1. Then an orthogonal matrix Q of eigenvectors of E can be chosen as*

$$Q = \begin{bmatrix} X & -Y \\ Y & X \end{bmatrix}, \quad X, Y \in \mathbb{R}^{2\times 2}, \quad X^T Y = Y^T X, \quad X^T X + Y^T Y = I_2.$$

Proof. Just take the first two columns of Q to be orthogonal eigenvectors relative to κ_1, κ_2. □

The next result is the key to the final simplified form with which we work and it says that the matrix Q of Corollary 3.2 tranforms M in a form that is like M itself.

Lemma 3.3. *Let E be as in Lemma 3.1, let Q be the matrix of Corollary 3.2 which diagonalizes E: $Q^T E Q = \begin{bmatrix} \kappa_1 & 0 & 0 & 0 \\ 0 & \kappa_2 & 0 & 0 \\ 0 & 0 & -\kappa_1 & 0 \\ 0 & 0 & 0 & -\kappa_2 \end{bmatrix}$, and let M be the function in* (3.1)*: $M =$*

$\left[\begin{smallmatrix} x & y & 0 & -z \\ y & -x & z & 0 \\ 0 & z & x & y \\ -z & 0 & y & -x \end{smallmatrix}\right]$. *Then, Q^TMQ has the form*

$$Q^TMQ = \begin{bmatrix} \chi & \eta & 0 & -\zeta \\ \eta & -\chi & \zeta & 0 \\ 0 & \zeta & \chi & \eta \\ -\zeta & 0 & \eta & -\chi \end{bmatrix} . \tag{3.2}$$

In other words, Q^TMQ has the same form of M for the new variables (χ, η, ζ).

Proof. An explicit computation with the given forms of Q and M gives for the blocks of Q^TMQ:

$$\begin{aligned} (Q^TMQ)_{11} &= X^TM_1X + X^TM_2Y - Y^TM_2X + Y^TM_1Y\,, \\ (Q^TMQ)_{22} &= Y^TM_1Y - Y^TM_2X + X^TM_2Y + X^TM_1X\,, \\ (Q^TMQ)_{12} &= -X^TM_1Y + X^TM_2X + Y^TM_2Y + Y^TM_1X\,, \\ (Q^TMQ)_{21} &= X^TM_1Y - X^TM_2X - Y^TM_2Y - Y^TM_1X\,. \end{aligned}$$

From these, recalling that $M_1^T = M_1$ and $M_2^T = -M_2$, we get that $(Q^TMQ)_{12} = -[(Q^TMQ)_{12}]^T = [(Q^TMQ)_{21}]^T = -(Q^TMQ)_{21}$ and thus $(Q^TMQ)_{12} = \left[\begin{smallmatrix} 0 & -\zeta \\ \zeta & 0 \end{smallmatrix}\right]$. Also, we have that $(Q^TMQ)_{11} = [(Q^TMQ)_{11}]^T = (Q^TMQ)_{22}$ from which it follows that $(Q^TMQ)_{11} = \left[\begin{smallmatrix} \chi & \eta \\ \eta & \xi \end{smallmatrix}\right]$. The final fact, that $\xi = -\chi$, is a consequence of the fact that M has zero trace, and thus so does Q^TMQ. □

Remark 3.4. We point out that, within the vector space of symmetric matrices E of the form considered in Lemma 3.1, those having distinct eigenvalues form an open and dense set. Openness is clear, since, if a matrix E as in Lemma 3.1 has distinct eigenvalues, continuity of the eigenvalues with respect to the entries of the matrix will ensure that all sufficiently close matrices will also have distinct eigenvalues. To validate density, suppose that a matrix E has a pair of coalescing eigenvalues, and consider its matrix Q of eigenvectors as given in Corollary 3.2. Then, we can choose $\varepsilon_1, \varepsilon_2 > 0$, and arbitrarily small, so that

$$E + Q \begin{bmatrix} \varepsilon_1 & 0 & 0 & 0 \\ 0 & \varepsilon_2 & 0 & 0 \\ 0 & 0 & -\varepsilon_1 & 0 \\ 0 & 0 & 0 & -\varepsilon_2 \end{bmatrix} Q^T$$

has the same structure of E and has distinct eigenvalues.

Summary 3.5. In light of Corollary 3.2 and Lemma 3.3, we can therefore consider the following structure for the function M and the matrix E (cfr. with Summary 2.5):

$$M = \begin{bmatrix} x & y & 0 & -z \\ y & -x & z & 0 \\ 0 & z & x & y \\ -z & 0 & y & -x \end{bmatrix} , \quad E = \begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & -a & 0 \\ 0 & 0 & 0 & -b \end{bmatrix} , \ a \geq b \geq 0\,, \tag{3.3}$$

where (x, y, z) are $\mathcal{C}^k$ functions of (α, β, γ) in a neighborhood of the origin, vanish at the origin, and can be used as a local coordinates system.

Here below we give results on the locus of points where $\lambda_2 = \lambda_3$, and where $\lambda_1 = \lambda_2$ and $\lambda_3 = \lambda_4$, in terms of the coordinate system (x, y, z). We will show that –under generic conditions on the coefficients of E– the set $\{(x, y, z) : \lambda_2 = \lambda_3\}$ is an ellipse (the green curve), lying in a certain plane. Moreover, the sets $\{(x, y, z) : \lambda_1 = \lambda_2\}$ and $\{(x, y, z) : \lambda_3 = \lambda_4\}$ (the blue and red curves) will be shown to be branches of hyperbola, lying in a plane perpendicular to that of the ellipse. Naturally, these are not exactly planes when viewed in the original (α, β, γ) coordinates.

3.1.1. *The Ellipse.* Next, we look at the set of points where $\lambda_2 = \lambda_3$.

Theorem 3.6. *Let M and E be as in* (3.3) *with $b > 0$, and let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ be the ordered eigenvalues of $M + \varepsilon E$, $\varepsilon > 0$. Then, when $z = 0$, $\lambda_2 = \lambda_3$ along the ellipse*

$$\left(\frac{2x}{a+b}\right)^2 + \left(\frac{y}{\sqrt{ab}}\right)^2 = \varepsilon^2 \,. \tag{3.4}$$

Proof. With abuse of notation, below let $a = \varepsilon a$ and $b = \varepsilon b$.

In this case of $z = 0$, the problem decouples in two subproblems corresponding to the diagonal blocks, with characteristic polynomials respectively given by

$$P_1(\lambda) := (x + a - \lambda)(-x + b - \lambda) - y^2 \qquad \text{and} \qquad P_2(\lambda) := (-x + a + \lambda)(x + b + \lambda) - y^2 \,.$$

Since $a \geq b > 0$, we have that λ_1 is a root of P_1 and λ_4 is a root of P_2. Therefore, $\lambda_2 = \lambda_3 \equiv \mu$ means that μ must satisfy both $P_1(\mu) = 0$ and $P_2(\mu) = 0$, that is $P_1(\mu) - P_2(\mu) = P_1(\mu) + P_2(\mu) = 0$. These two relations give the system

$$\begin{cases} \mu^2 + ab - (x^2 + y^2) = 0 \\ \mu(a + b) + (a - b)x = 0 \end{cases} . \tag{3.5}$$

Solving for μ from the second equation and substituting in the first gives the relation

$$ab = x^2\left(1 - \frac{(b-a)^2}{(a+b)^2}\right) + y^2 \ ,$$

which is nothing but (3.4). □

Remark 3.7. In Theorem 3.6, we have assumed that in E we have $b > 0$. Naturally, this is a generic condition; indeed, it corresponds to saying that E does not have a double eigenvalue at 0. But, even if this condition is violated (that is, $b = 0$), it is a simple computation from (3.5) to observe that the ellipse degenerates in the line given by the x-axis. Finally, if we have at once $a = b = 0$ this means that there is no perturbation, and this situation is obviously of no interest.

3.1.2. *The Hyperbola.* Next, and still for the problem (3.3), we look at the red and blue curves.

Theorem 3.8. *Let M and E be as in* (3.3)*, with $a > b$. Let $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4$ be the ordered eigenvalues of $M + \varepsilon E$, $\varepsilon > 0$. Let $\sigma = \varepsilon(a+b)/2$ and $\delta = \varepsilon(a-b)/2$. Then, for $y = 0$, we have $\lambda_1 = \lambda_2$ and $\lambda_3 = \lambda_4$ occurring along one distinct branch of the hyperbola characterized as*

$$\left(\sqrt{(\sigma+x)^2+z^2} - \sqrt{(\sigma-x)^2+z^2}\right)^2 - 4\delta^2 = 0\,. \tag{3.6}$$

Proof. We consider the eigenvalues of $M(x,0,z) + \varepsilon E$ (that is, we set $y = 0$). We have:

$$\lambda_1, \lambda_2 = \pm\,\delta + \sqrt{(\sigma \pm x)^2 + z^2}\,,$$
$$\lambda_3, \lambda_4 = \pm\,\delta - \sqrt{(\sigma \pm x)^2 + z^2}\,.$$

Now, $\lambda_i \geq \lambda_j$, for all $i \in \{1,2\}$ and $j \in \{3,4\}$, follows from

$$\sqrt{(\sigma-x)^2+z^2} + \sqrt{(\sigma+x)^2+z^2} \geq |\sigma - x| + |\sigma + x| \geq 2\sigma \geq 2\delta\,,$$

and three trivial inequalities. Finally, we simply observe that

$$(\lambda_1 - \lambda_2)(\lambda_3 - \lambda_4) = 0$$

is equivalent to (3.6). In particular, we note that $\lambda_1 = \lambda_2$ occurs for $x < 0$, while $\lambda_3 = \lambda_4$ occurs for $x > 0$. This completes the proof. □

Remarks 3.9.

(i) We can rewrite (3.6) as

$$\left(\sqrt{(\sqrt{\rho^2+\tau^2}+x)^2+z^2} - \sqrt{(\sqrt{\rho^2+\tau^2}-x)^2+z^2}\right)^2 - 4\rho^2 = 0\,, \tag{3.7}$$

with $\rho = \varepsilon(a-b)/2$ and $\tau = \varepsilon\sqrt{ab}$, which makes us recognize (3.6) as the set of points whose difference between the distance to the foci is kept constant and equal to 2ρ. Here, the foci are $(-\sqrt{\rho^2+\tau^2}, 0)$ and $(\sqrt{\rho^2+\tau^2}, 0)$, and the eccentricity of the hyperbola is $\sqrt{\tau^2/\rho^2+1}$. As a consequence, we have the canonical form of the hyperbola simply as (if $b > 0$)

$$\frac{x^2}{\rho^2} - \frac{z^2}{\tau^2} = 1\,, \quad \rho = \varepsilon(a-b)/2\,,\ \tau = \varepsilon\sqrt{ab}\,. \tag{3.8}$$

(ii) Each branch of the hyperbola (3.8) pierces the (x,y)-plane on the x-axis at the points $x = \pm\varepsilon\frac{a-b}{2}$, which are inside the ellipse (3.4) for $a > b$.

(iii) For completeness, we remark that if $a = b$, then the two branches of the hyperbola degenerate into the z-axis, obviously passing inside the ellipse (3.4), which in this case is a circle, just at the origin.

The next result shows that, except for the ellipse (3.4) and the two branches of the hyperbola (3.8), there are no more coalescing points for $\lambda_i = \lambda_{i+1}$, $i = 1,2,3$.

Theorem 3.10. *Let M and E be as in* (3.3)*, with $a>0$. If $y\neq 0$ and $z\neq 0$, then all the eigenvalues of $M+\varepsilon E$, $\varepsilon>0$, are simple.*

Proof. Let $\lambda\in\mathbb{R}$, and consider

$$M+\varepsilon E-\lambda I=\begin{bmatrix} N_{11} & N_{12}\\ N_{12}^T & N_{22}\end{bmatrix},$$

where

$$N_{11}=\begin{bmatrix} x+\varepsilon a-\lambda & y\\ y & -x+\varepsilon b-\lambda\end{bmatrix},\ N_{12}=\begin{bmatrix} 0 & z\\ -z & 0\end{bmatrix},\ N_{22}=\begin{bmatrix} x-\varepsilon a-\lambda & y\\ y & -x-\varepsilon b-\lambda\end{bmatrix}.$$

Assume $z\neq 0$. Then N_{12} is invertible, and $\operatorname{rank}(M+\varepsilon E-\lambda I)\geq 2$.

Moreover

$$\operatorname{rank}(M+\varepsilon E-\lambda I)=\operatorname{rank}\left(\begin{bmatrix} N_{12} & N_{11}\\ N_{22} & N_{12}^T\end{bmatrix}\right)=\operatorname{rank}\left(\begin{bmatrix} N_{12} & N_{11}\\ N_{22} & N_{12}^T\end{bmatrix}\begin{bmatrix} I & -N_{12}^{-1}N_{11}\\ 0 & I\end{bmatrix}\right)=$$

$$=\operatorname{rank}\left(\begin{bmatrix} N_{12} & 0\\ N_{22} & N_{12}^T-N_{22}N_{12}^{-1}N_{11}\end{bmatrix}\right).$$

Direct computation yields

$$(N_{12}^T-N_{22}N_{12}^{-1}N_{11})_{11}=\frac{2\varepsilon a y}{z}.$$

Therefore, if $y\neq 0$, then $\operatorname{rank}(M+\varepsilon E-\lambda I)\geq 3$. □

We summarize in the following Theorem, which validates Figure 1 in this simpler "constant E" case.

Theorem 3.11. *Let M and E be as in* (3.1)*, with E constant and with distinct eigenvalues. Then, locally, the eigenvalue* 0 *at the origin of* (2.3)*, of multiplicity* 4*, splits into four eigenvalues $\lambda_1\geq\lambda_2\geq\lambda_3\geq\lambda_4$ of $M+\varepsilon E$, $\varepsilon>0$, that are distinct except as follows:*

(i) *$\lambda_1=\lambda_2$ and $\lambda_3=\lambda_4$ along two branches of the same hyperbola;*
(ii) *$\lambda_2=\lambda_3$ along an ellipse;*
(iii) *the hyperbola and ellipse of above points (i) and (ii) lie in two perpendicular planes with the two branches of the hyperbola passing inside the ellipse.*

□

3.2. **The general problem.** Now we consider the full problem, given by (2.5), repeated below for convenience:

$$\begin{bmatrix} M_1+\varepsilon\widehat{E} & \varepsilon F\\ \varepsilon F^T & M_2+\varepsilon H\end{bmatrix}, \tag{3.9}$$

where M_1 is as in (2.3) and $\widehat{E}$ has the structure given in Lemma 2.9, and both are $\mathcal{C}^k$ functions of α,β,γ (in a neighborhood of the origin).

What we are going to do is to validate all the steps of the analysis of Section 3.1, from which the end result will follow. First of all, we observe that, by virtue of Theorem 2.10, we can focus our attention on the block $\widetilde{M_1}=\widetilde{M}+\varepsilon\widetilde{E}$, with $\widetilde{M}$ and $\widetilde{E}$ as in (2.7).

Before proceeding, we make the following assumption, which is generic in light of Remark 3.4.

Assumption 3.12. At $(\alpha,\beta,\gamma)=(0,0,0)$, $\widehat{E}$ has distinct eigenvalues, which therefore will remain distinct in an open ball B_0 centered at the origin.

Below we give the steps that conduce to the sought result:

(a) "*On* $\widetilde{M}$".

(i) Given the form of $\widetilde{M}$ in (2.7), we can shift the problem by $(\widetilde{\alpha}+\widetilde{\delta})/2$, so that we can assume to have the following form for $\widetilde{M}$

$$\widetilde{M}=\begin{bmatrix} \widetilde{\alpha} & \widetilde{\beta} & 0 & -\widetilde{\gamma} \\ \widetilde{\beta} & -\widetilde{\alpha} & \widetilde{\gamma} & 0 \\ 0 & \widetilde{\gamma} & \widetilde{\alpha} & \widetilde{\beta} \\ -\widetilde{\gamma} & 0 & \widetilde{\beta} & \widetilde{-\alpha} \end{bmatrix}. \tag{3.10}$$

(ii) Since $\alpha=\beta=\gamma=0$ is a generic CI point, then the Jacobian

$$J(\varepsilon)=\begin{bmatrix} \partial_\alpha\widetilde{\alpha} & \partial_\beta\widetilde{\alpha} & \partial_\gamma\widetilde{\alpha} \\ \partial_\alpha\widetilde{\beta} & \partial_\beta\widetilde{\beta} & \partial_\gamma\widetilde{\beta} \\ \partial_\alpha\widetilde{\gamma} & \partial_\beta\widetilde{\gamma} & \partial_\gamma\widetilde{\gamma} \end{bmatrix}_{(\alpha,\beta,\gamma)=(0,0,0)}$$

is invertible for ε sufficiently small.

(iii) It follows that we can change variables and use

$$x(\varepsilon)=\widetilde{\alpha}\ ,\quad y(\varepsilon)=\widetilde{\beta}\ ,\quad z(\varepsilon)=\widetilde{\gamma}\,,$$

for ε sufficiently small and in a neighborhood of $(\alpha,\beta,\gamma)=(0,0,0)$.

In short, we can assume that we are dealing with the form (cfr. with (2.3))

$$\widetilde{M}=\begin{bmatrix} x & y & 0 & -z \\ y & -x & z & 0 \\ 0 & z & x & y \\ -z & 0 & y & -x \end{bmatrix}, \tag{3.11}$$

where the functions x,y,z, are $\mathcal{C}^\omega$ in ε, $\mathcal{C}^k$ in α,β,γ, vanish at the origin $(\alpha,\beta,\gamma)=(0,0,0)$ for $\varepsilon=0$, and can be used as local coordinates.

(b) "*On* $\widetilde{E}$". Here we look at the term $\widetilde{E}$ in (2.7):

$$\widetilde{E}=\varepsilon(\widehat{E}+\sum_{k=1}^{\infty}\varepsilon^{2k}E_{2k+1})=\begin{bmatrix} \widetilde{E}_1 & \widetilde{E}_2 \\ \widetilde{E}_2 & -\widetilde{E}_1 \end{bmatrix}\ ,\quad \widetilde{E}_1=\begin{bmatrix} \widetilde{a} & \widetilde{b} \\ \widetilde{b} & \widetilde{c} \end{bmatrix}\ ,\ \widetilde{E}_2=\begin{bmatrix} \widetilde{d} & \widetilde{e} \\ \widetilde{e} & \widetilde{f} \end{bmatrix}\ .$$

As in Lemma 3.1, the eigenvalues of $\widetilde{E}$ are still of the type $\pm\kappa_1$ and $\pm\kappa_2$. Moreover, under Assumption 3.12 we have that at $(\alpha,\beta,\gamma)=(0,0,0)$, $\widehat{E}$ has distinct eigenvalues, and therefore $\widetilde{E}$ has distinct eigenvalues in an open ball B_0 centered at $(\alpha,\beta,\gamma)=(0,0,0)$ and for ε sufficiently small. Since $\widetilde{E}$ is analytic in ε, $\mathcal{C}^k$ in (α,β,γ), and symmetric, its eigendecomposition around the origin is analytic in ε

and $\mathcal{C}^k$ in (α,β,γ). As a consequence, Corollary 3.2 and Lemma 3.3 still hold (locally), and we can therefore still consider the following structure for the functions $\widetilde{M}$ and $\widetilde{E}$:

$$\widetilde{M}=\begin{bmatrix} x & y & 0 & -z \\ y & -x & z & 0 \\ 0 & z & x & y \\ -z & 0 & y & -x \end{bmatrix}\,,\quad \widetilde{E}=\begin{bmatrix} a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & -a & 0 \\ 0 & 0 & 0 & -b \end{bmatrix}\,,\ a>b>0\,, \tag{3.12}$$

for (α,β,γ) in a sufficiently small neighborhood of the origin and for ε in a sufficiently small interval around 0.

(c) "*On the blue, red, green, curves*". At this point, and with the caveat of the necessity to appropriately restrict ourselves to (α,β,γ) in a sufficiently small neighborhood of the origin and for ε in a sufficiently small interval around 0, the arguments of Section 3.1 on the ellipse and hyperbola still hold, much like they did there.

In an admittedly emphatic way, we can summarize the end result in Theorem 3.13 below.

Theorem 3.13. *Figure 1 is qualitatively correct.* □

4. Algorithm and an Example

Our ultimate goal is to detect configurations such as the one depicted in Figure 1, which betray the presence of a generic coalescing point for a complex Hermitian problem (and we reiterate that we are exclusively concerned with the generic case, see Assumption 2.4 and Definition 2.2). To this end, we will consider computing curves of coalescing points for a generic symmetric function $\widehat{A}$ of size $2n$, depending on three real parameters. Our algorithm will do two things: compute a curve of coalescing points and detect the presence of other nearby curves.

4.1. **Computation of curves of coalescing points.** For the computation of curves of coalescing points, we propose a method that follows closely the idea of predictor-corrector path following algorithms, see for instance [13]. Consider the set $\Gamma_j=\{\xi\,:\,\lambda_j(\xi)=\lambda_{j+1}(\xi)\}$, for some $0\le j\le 2n-1$.

As it is well understood (e.g., see [5, 10]), Γ_j is -locally– a smooth manifold embedded in $\mathbb{R}^3$ (i.e., a curve). In general, the set Γ_j will be made of several connected components (curves), called branches. Here we address the problem of how to obtain a portion of one branch γ of Γ_j that lies inside a given box $\Omega\subset\mathbb{R}^3$. The computed branch will consist of consecutive (with respect to arc length) points $\xi^{(0)},\xi^{(1)},\dots$ on γ.

Below we see how, given $\xi^{(0)},\dots,\xi^{(k)}$, we obtain the next point $\xi^{(k+1)}$. It is essentially a three-stage process: predictor, corrector and acceptance/rejection of the step.

4.1.1. *Predictor.* First, we form the predictor $\xi^{(k+1)}_{\rm pred}$ by taking a step of length h (see below) along an approximate tangent direction T to the curve at $\xi^{(k)}$. That is, see Figure 2, we set

$$\xi^{(k+1)}_{\rm pred} = \xi^{(k)} + h\,T\,,$$

with the following choices.

1. If $k \geq 2$, the vector T is the unit tangent vector at $\xi^{(k)}$ to the curve obtained by quadratic interpolation at the points $\xi^{(k-2)}, \xi^{(k-1)}$ and $\xi^{(k)}$, chosen so that it has positive component in the direction of $\xi^{(k)} - \xi^{(k-1)}$; for $k = 0$, we set $T = 0$ (trivial predictor), while, for $k = 1$, we take T to be the unit vector in the direction of $\xi^{(1)} - \xi^{(0)}$ (secant predictor).
2. The step length h is chosen adaptively as follows. Let $h_{\rm old}$ be the step length used to compute $\xi^{(k)}$. We set

$$h = \tau\, h_{\rm old}, \qquad \text{where } \tau = \sqrt{\frac{\texttt{tol}}{\left\|\xi^{(k)}_{\rm pred} - \xi^{(k)}\right\|}}\,, \tag{4.1}$$

which aims to achieve $\left\|\xi^{(k)}_{\rm pred} - \xi^{(k)}\right\| \approx \texttt{tol}$ for all k (note that $\left\|\xi^{(k)}_{\rm pred} - \xi^{(k)}\right\|$ is expected to be of order $\mathcal{O}(h^2)$). We also always enforce $h_{\rm min} \leq h \leq h_{\rm max}$; see Section 4.4 for our choices of $h_{\rm min}$, $h_{\rm max}$.

4.1.2. *Corrector.* The predictor $\xi^{(k+1)}_{\rm pred}$ is refined by searching for the point

$$\xi_{\rm corr} \in \arg\min_{\xi\in\Pi} F(\xi) = (\lambda_j(\xi) - \lambda_{j+1}(\xi))^2$$

which is closest to $\xi^{(k+1)}_{\rm pred}$, where Π is the plane through the predictor and perpendicular to T. The minimization problem is solved by seeking a stationary point for F through Newton's method, where the predictor serves as initial guess. If convergence of Newton's iterations fails, a new predictor is formed by halving the step length h, until convergence is successful or h falls below $h_{\rm min}$. (See also [4, Section 3.3]).

4.1.3. *Step acceptance/rejection.* If the corrector stage was completed successfully, we update the value of τ using again the formula given in (4.1), with $\xi^{(k+1)}_{\rm pred}$ and $\xi_{\rm corr}$ in place of, respectively, $\xi^{(k)}_{\rm pred}$ and $\xi^{(k)}$. Presently, we will always enforce that $\tau \geq \tau_{\rm min}$, where in all of our experiments we have successfully used $\tau_{\rm min} = 0.7$. Thus, if $\tau < \tau_{\rm min}$, the point $\xi_{\rm corr}$ is rejected, the predictor $\xi^{(k+1)}_{\rm pred}$ is recomputed with step length reduced by the (updated) factor τ and the predictor-corrector stages are repeated. Otherwise, we set $\xi^{(k+1)} = \xi_{\rm corr}$, and consider the step to be successfully completed. Note that enforcing this rejection strategy effectively limits the maximum distance between the predictor and the computed approximation, i.e. for all k we must have

$$\left\|\xi^{(k)}_{\rm pred} - \xi^{(k)}\right\| \leq \frac{\texttt{tol}}{\tau^2_{\rm min}}\,. \tag{4.2}$$

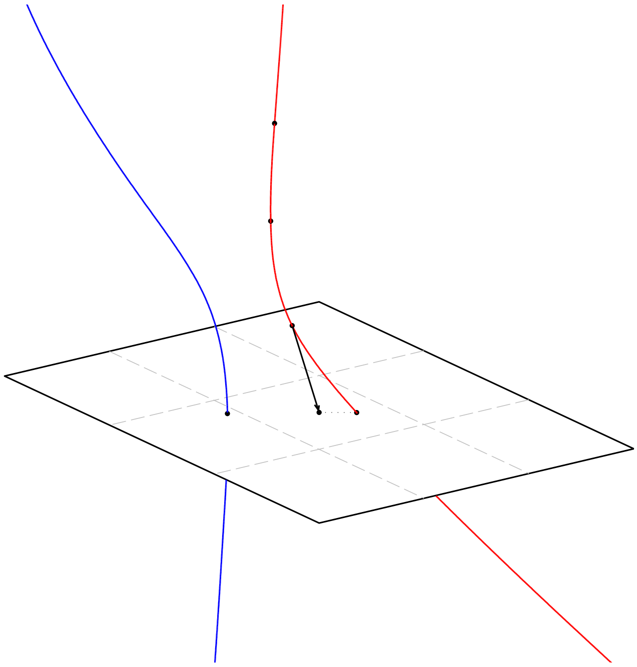

FIGURE 2. Computation of a curve of coalescing points, and search for nearby curves.

The continuation algorithm progresses as described above until either $h < h_{\min}$ or the branch under consideration has been completely approximated. For the latter case, we should expect one of the following two situations to occur: (i) we have followed the branch, moving away from $\xi^{(0)}$ in both directions, until we have stepped out of Ω, (ii) the branch is a closed curve that lies completely inside Ω. While the first scenario is very easy to detect, the second one needs to be handled with care. Below we give a description of the method we have implemented to detect when situation (ii) occurs.

4.2. **Detection of closed curves.** Detecting when a branch is closed is a critical task of a path-following algorithm, as failing to do so will cause the algorithm to enter an infinite loop. This is even more critical in our context, where closed curves are a typical (so much as desired) feature that we want to correctly identify. The method we have implemented is an adaptation of the *piercing computation* proposed in [17] in the context of continuation of implicitly defined two-dimensional manifolds.

We want to detect when two consecutive points $\xi^{(k)}$ and $\xi^{(k+1)}$ on γ "bracket" the initial point $\xi^{(0)}$ (see Figure 3). To do so, at the start of the process we define the piercing plane

$\Pi_{\rm p}$ as the plane through $\xi^{(0)}$ perpendicular to the vector $\xi^{(1)} - \xi^{(0)}$. Upon computation of each new point $\xi^{(k+1)}$ on γ, we check whether $\xi^{(k)}$ and $\xi^{(k+1)}$ fall on opposite sides of $\Pi_{\rm p}$. If so, we compute the intersection of the line segment that joins $\xi^{(k)}$ to $\xi^{(k+1)}$ with the plane $\Pi_{\rm p}$, and then "correct" it along $\Pi_{\rm p}$ similarly to what is done in Section 4.1.2. If the corrected point is within a given distance δ from $\xi^{(0)}$, we declare the branch γ to be closed, and terminate the computation. (Only even crossings of $\Pi_{\rm p}$ need to be examined, of course).

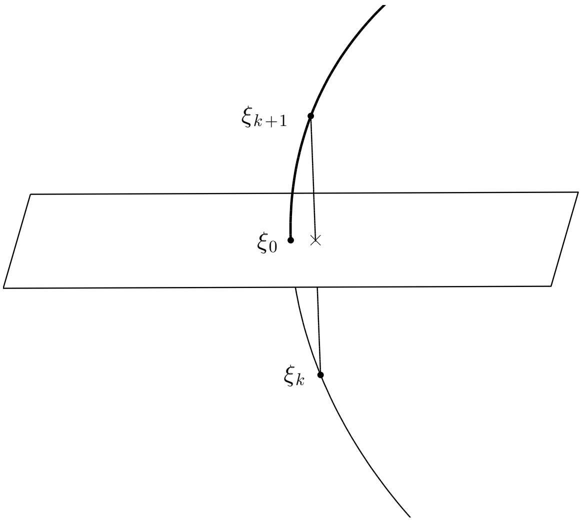

Figure 3. Detection of closed curves.

4.3. **Detection of nearby curves.** While we compute a branch of coalescing points of Γ_j, we also want to detect other branches of Γ_q, $q \neq j$, that pass nearby. To do so, we add an additional module to the overall algorithm. With the same notation introduced in Section 4.1, suppose $\xi^{(k+1)}$ has just been successfully computed. Before moving to the next step, we consider a square S of side length l centered at $\xi^{(k+1)}_{\rm pred}$ on the plane Π. We subdivide S through a regular grid, and swipe the grid in search of coalescing points of any possible pair of eigenvalues of $\widehat{A}$. Restricting $\widehat{A}$ to S, this search is done with the method developed in [6] for the computation of coalescing points of real symmetric matrix functions of two parameters. See Figure 2. If coalescing points, other then $\xi^{(k+1)}$, are detected, they are saved in a database, and will serve as initial seeds for the computation of more curves of coalescing points.

4.4. **Example.** Below, we illustrate the performance of the previous algorithm, and further show how the computed curves are used to locate coalescing points of the original complex Hermitian problem.

Consider the function A, for $(\alpha, \beta, \gamma) \in [0,1]^3$:

$$A(\alpha,\beta,\gamma) = (1-\frac{1}{2}\alpha^2)H_1 + \alpha H_2 + (1-\frac{1}{2}\beta^2)H_3 + \beta H_4 + (1-\frac{1}{2}\gamma^2)H_5 + \gamma H_6\,,$$

where $H_1,\dots,H_6$ are 6×6 complex Hermitian matrices explicitly given in [4, Example 4.1].

With the method developed in [4], the following three coalescing points were detected and approximated inside the cube $[0,1]^3$:

$$\begin{aligned} &(i) \quad \mu_1 = \mu_2 \ \text{ at } \ \xi_1 = \begin{bmatrix}0.44511899 & 0.34014156 & 0.94489258\end{bmatrix}\ ;\\ &(ii) \quad \mu_2 = \mu_3 \ \text{ at } \ \xi_2 = \begin{bmatrix}0.46761305 & 0.46167575 & 0.44946999\end{bmatrix}\ ;\\ &(iii) \quad \mu_5 = \mu_6 \ \text{ at } \ \xi_3 = \begin{bmatrix}0.80644491 & 0.87260280 & 0.41732847\end{bmatrix}\ . \end{aligned} \tag{4.3}$$

Below, we show how the approach proposed in this paper is used to locate the coalescing points of A in $\Omega = [-0.25, 1.25]^3$.

First, we form $M_{\text{pert}} = M + \varepsilon E$ as in (1.2). This M_{pert} plays the role of $\widehat{A}$ in Section 4. The entries of E are chosen as independent samples from the uniform distribution in $[-1,\ 1]$ and $\varepsilon > 0$ is chosen so that $\|\varepsilon E\| = 10^{-1}$ (we note that, because of Lemma 2.8, the results would be identical –for the same E– by choosing $\varepsilon < 0$). Note that M_{pert} is a 12×12 real symmetric matrix function.

The next goal is to compute branches of the curves of coalescing points $\Gamma_1, \dots, \Gamma_{11}$ for A inside Ω. We recall that our notation is such that along Γ_j, the j-th and $(j+1)$-th eigenvalue coalesce. Computation of these curves is done as follows.

- First, we look for the initial points needed to start the computation by searching on each of the faces of the cube $[0,1]^3$ for coalescing points of M_{pert}; this computation is done similarly to what we described in Section 4.3. We find 20 initial points, and place them into a queue $\mathcal{Q}$.
- Then, we repeatedly pick a point in $\mathcal{Q}$ and compute the portion of the corresponding branch that lies in Ω. For the continuation, we chose $h_{\min} = 10^{-14}$, $h_{\max} = 10^{-2}$ and $\texttt{tol} = 10^{-3}\,\tau_{\min}^2$. (Recall that $\tau_{\min} = 0.7$). Once a branch has been successfully computed, it is added to a database $\mathcal{B}$ of all computed branches.
- While continuing each branch, we monitor the presence of other nearby branches as described in Section 4.3, with the square S there having side of length $l = 0.1$, and further subdivided into 3×3 equal sub-squares (see Figure 2). If a point on a (potentially) new branch is detected, it is added to $\mathcal{Q}$.
- Finally, before starting computation of a branch from a point η in $\mathcal{Q}$, and $\eta \in \Gamma_j$ for some j, we make sure that it is not part of a branch of Γ_j that we already computed (and hence in $\mathcal{B}$). To this end, we run a *proximity test*: we check whether, for some ξ on a branch of Γ_j in $\mathcal{B}$, we have $\|\eta - \xi\| \le d/2$, where

$$d = \sqrt{h_{\max}^2 + \left(\frac{\texttt{tol}}{\tau_{\min}^2}\right)^2} \tag{4.4}$$

is the maximum distance between two consecutive points on any computed branch; for us, $d \approx 1.005 \times 10^{-2}$. If the condition above is satisfied, the point η is removed from $\mathcal{Q}$. The process continues until the queue $\mathcal{Q}$ is empty.

Remark 4.1. Of course, it is critical to choose carefully the parameters involved in the computation. Our choices have been dictated by the fact that we expect the size of the closed branches of coalescing points for M_{pert}, that originate close to a coalescing point of A, to be $\mathcal{O}(\varepsilon\,\|E\|)$.

(i) To properly detect when a branch is closed, the distance δ of Section 4.3 has to be taken orders of magnitude smaller then the expected diameter of the closed curves, but safely away from the tolerance used to declare convergence of Newton's method. For this reason, in all our experiments we have chosen $\delta = 10^{-7}$.

(ii) We need to choose l larger than d in order to possibly discover new branches while avoiding "false positives" (i.e. erroneously discard a new branch) during proximity checks.

In Figure 4, we show the 13 branches of coalescing eigenvalues we have computed. We found branches from the curves Γ_j, for $j = 1, 2, 3, 4, 5, 7, 9, 10, 11$. All but Γ_7 are related to the coalescing points ξ_1, ξ_2 and ξ_3 of (4.3). Note that there are four configurations similar to the one of Figure 1. Of these four, three are related to the points in (4.3), and the fourth (top one) reveals a coalescing point just outside the cube $[0,1]^3$.

To complete the experiment, we compute the centroid of each of the small closed branches found for Γ_{2j}, and refine it through Newton's method (similar to what we described in Section 4.1.2). Newton's method always converged in a few iterations, and we ended up correctly approximating the three coalescing points of (4.3) and found a new one outside of $[0,1]^3$:

$$\mu_2 = \mu_3 \text{ at } \xi_4 = \begin{bmatrix} 0.79393735 & 0.73364539 & 1.0768356 \end{bmatrix} .$$

Remark 4.2. Finally, we point out that –based on our experience for the present example– the algorithm we proposed in this work has proven to be quite robust, and it never failed either to compute a branch (i.e. $h < h_{\min}$ never occurred in our experiments) or to converge to a conical intersection of the original Hermitian problem.

5. Conclusions

In this work we considered perturbation of conical intersections (CIs) of a Hermitian function $A = B + iC \in \mathbb{C}^{n\times n}$ depending on three real parameters, by studying generic symmetric perturbations $M_{\text{pert}} = M + \varepsilon E$ of the associated symmetric function $M = \left[\begin{smallmatrix} B & -C \\ C & B \end{smallmatrix}\right]$. For M_{pert}, the eigenvalues now coalesce along curves and we gave rigorous results on the local structure of these curves; in a nutshell, we validated the qualitative features of Figure 1, and proved (at all order in ε) that –near a CI of the original A– there are three such curves, a nearly ellipsoidal closed curve (the green curve), containing the original CI, and two other curves (the red and blue curves) which are nearly branches of hyperbola passing inside the green curve. We further proposed and implemented an algorithm which

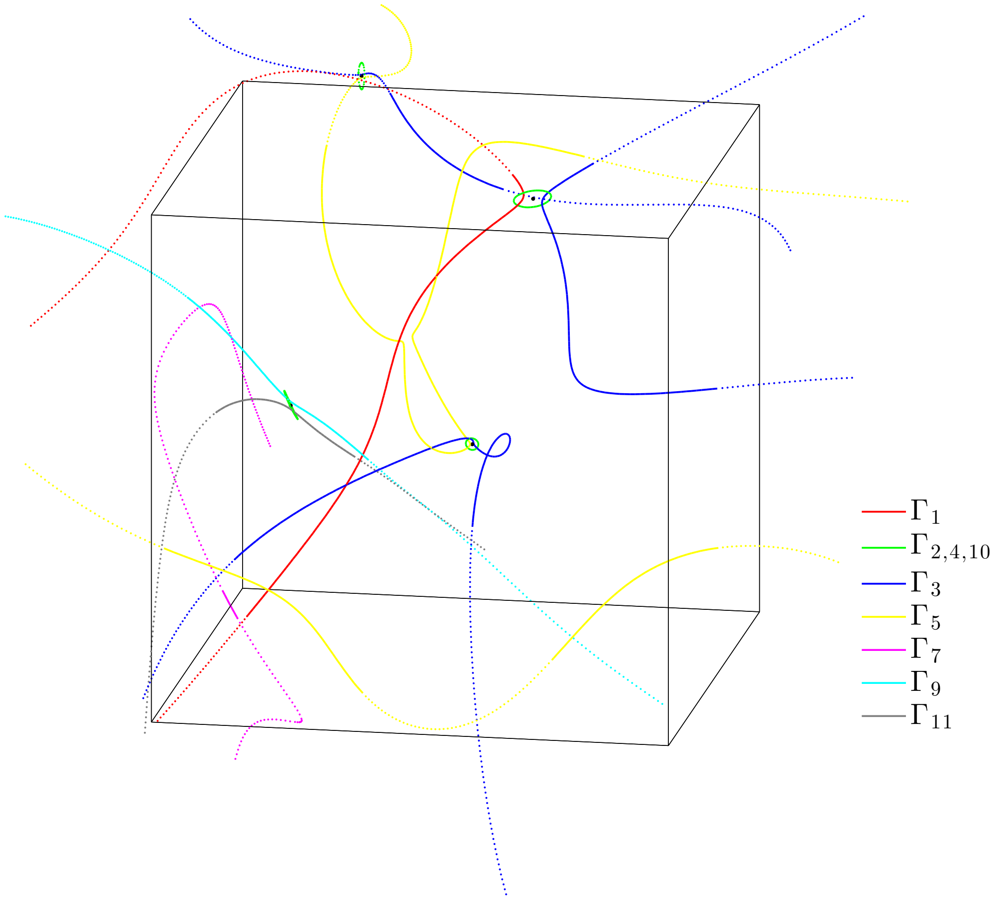

Figure 4. Example 4.4: Curves of coalescing points for enlarged and perturbed symmetric problem M_{pert}. Curves are "dotted" outside of the cube $[0,1]^3$. The four CIs of the original Hermitian problem are indicated by a black dot.

computes (globally) these curves (red, blue, and green), and finally showed how from these we can approximate CI points of the function A.

From the theoretical point of view, our work complements that of [14, 15], and from the algorithmic/computational point of view, it complements that of [4].

In [14, 15], the authors unfold the singularity given by the CI by perturbing the Hermitian problem with a non-Hermitian perturbation, and show that –at leading order– the original CI gets replaced locally by a closed ring of "exceptional points", that is parameter values where the perturbed problem has a non-trivial Jordan block. In spite of the similar flavor of having the original CI being replaced by a closed curve, our approach and that of these cited works are fundamentally different. Most notably, our methodology leads to robust algorithmic development for locating the CIs. This is due to the key aspect of our development: we perturb the enlarged symmetric problem with a symmetric perturbation. Although our perturbed problem cannot be directly interpreted as a perturbation of the function A, the main advantage of our study is that not only the green curve becomes available, but –of equal importance– also the red and blue curves. Indeed, these red and blue curves are of paramount importance in the development of our computational method to locate the CI of A, since with our computational technique we create a skeleton of all red and blue (and green) curves from which we can approximate the CIs of the original problem.

The computational method we developed in this work differs from the purely topological method we recently studied in [4], and is complementary to that. A detailed comparison of these two methods remain to be done, but there are obvious benefits to either technique. For example, the technique of [4] provides rigorous enclosure regions for generic CIs. At the same time, the technique we examined in this work also presents the nontrivial advantages below.

(i) The expensive global 3-d search of [4] for CIs over a region Ω (say, a cube), gets now replaced by 2-d searches: over the boundary of Ω (faces of the cube) to locate starting points for the (red and blue) curves of coalescing eigenvalues of M_{pert}, and over small planar regions while continuing these curves.

(ii) The present technique allows to search for coalescing points for any selected pair of eigenvalues of A, unlike the technique of [4] which can only be adapted to approximate dominant (or nearly dominant) pairs. This is particularly convenient when the dimension n is large and we are interested in a CI for a pair of eigenvalues near "the middle of the spectrum."

(iii) The present method is particularly appealing when there are very few CIs of A, and the associated red/blue curves extend outside of our chosen search cube; this way, by looking for starting points on the faces of the cube, computing the red and blue curves will lead us to the CIs.

Finally, it should be possible to adapt our methodology and algorithm to the case of the SVD, as well as to the case of a Hermitian positive definite pencil. This also remains to be done in future work.

Appendix

Here we prove Theorem 2.10. The key ingredient in the proof is related to the solution of a Riccati equation; see Theorem A.1 below. In a similar context, this tool was used by Stewart (see [19]), and later extensively adopted by Stewart & Sun (see [20]) to obtain first order perturbation results on invariant subspaces and related eigenvalue problems. However, we want to take into consideration smooth dependence on parameters, and seek complete details –in our context– on the smoothness (with respect to parameters) and on the structure of all factors involved, as detailed in Theorem 2.10. Unfortunately, these cited results do not provide such refined detail (nor have we found it anywhere else), and we thus need to provide complete proofs.

Theorem A.1. *Let $(\alpha,\beta,\gamma)\in B$, an open ball centered at the origin of $\mathbb{R}^3$. Let $M_{11}\in\mathcal{C}^k(B,\mathbb{R}^{q\times q})$, $M_{22}\in\mathcal{C}^k(B,\mathbb{R}^{m\times m})$, $E_{12}\in\mathcal{C}^k(B,\mathbb{R}^{q\times m})$, $E_{21}\in\mathcal{C}^k(B,\mathbb{R}^{m\times q})$, $E_{11}\in\mathcal{C}^k(B,\mathbb{R}^{q\times q})$, $E_{22}\in\mathcal{C}^k(B,\mathbb{R}^{m\times m})$, for some $p\geq 0$. Also, let $\varepsilon\in\mathbb{R}$ be a real parameter.*

Consider the following nonlinear system for the unknown function X taking values in $\mathbb{R}^{m\times q}$:

$$\varepsilon E_{21}-X(M_{11}+\varepsilon E_{11})+(M_{22}+\varepsilon E_{22})X-\varepsilon XE_{12}X \;=\; 0\,. \tag{A.1}$$

Further, let $\|E_{ij}\|\leq\eta$, for all $(\alpha,\beta,\gamma)\in B$ and $i,j=1,2$, and assume that

$$\min_{1\leq i\leq q,\;1\leq j\leq m}|\lambda_i(M_{11})-\lambda_j(M_{22})|\geq\Delta>0,\qquad\forall(\alpha,\beta,\gamma)\in B\,. \tag{A.2}$$

Then, there exist an interval $I_0\equiv(-\varepsilon_0,\varepsilon_0)$ and an open ball $B_0\subseteq B$ centered at the origin of $\mathbb{R}^3$, such that (A.1) *has a unique solution X which is analytic in ε for $\varepsilon\in I_0$, and $\mathcal{C}^k$ in (α,β,γ) for $(\alpha,\beta,\gamma)\in B_0$, and such that $X=0$ for $\varepsilon=0$ and for any $(\alpha,\beta,\gamma)\in B_0$.*

For $\varepsilon\in I_0$, this unique solution X can be explicitly written as:

$$X=\sum_{k=1}^{\infty}\varepsilon^k X_k(\alpha,\beta,\gamma)\,, \tag{A.3}$$

where each $X_j\in\mathcal{C}^k(B_0,\mathbb{R}^{m\times q})$.

Proof. First of all, write the nonlinear system (A.1) more compactly as $F(X,\varepsilon,\alpha,\beta,\gamma)=0$, where F represents the left-hand-side of (A.1). Observe that F is $\mathcal{C}^\omega$ in X, and that for $\varepsilon=0$, the only solution of (A.1) is $X=0$, for any $(\alpha,\beta,\gamma)\in B$. Indeed, for $\varepsilon=0$, $F=0$ reduces to the linear system

$$XM_{11}-M_{22}X=0$$

which has the unique solution $X=0$ because of (A.2).

1. Now, since F is a $\mathcal{C}^k$ function of $(\varepsilon,\alpha,\beta,\gamma)\in I\times B$, and the derivative

$$F_X|_{\varepsilon=0}\;:\;Z\;\rightarrow\;-ZM_{11}+M_{22}Z$$

is invertible (because of (A.2)), then the implicit function theorem guarantees the existence of open neighborhoods R_0 of the origin of $\mathbb{R}^{m\times q}$, and of the origin of $I\times B$, say $I_0\times B_0$, where there is a unique $\mathcal{C}^k$ function $\widetilde{X}(\varepsilon,\alpha,\beta,\gamma)$ solution of $F=0$, with $\widetilde{X}\in R_0$ for $(\varepsilon,\alpha,\beta,\gamma)\in I_0\times B_0$.

2. Next, we look directly for a solution of (A.1) as power series expansion in ε. So, since $X=0$ for $\varepsilon=0$, we seek an expansion as in (A.3):

$$X=\varepsilon X_1+\varepsilon^2 X_2+\cdots\ .$$

Direct substitution of this expansion in (A.1) shows that:

$$X_1:\quad X_1M_{11}-M_{22}X_1=E_{21}\ ,$$

$$X_2:\quad X_2M_{11}-M_{22}X_2=E_{22}X_1-X_1E_{11}\ ,$$

$$X_3:\quad X_3M_{11}-M_{22}X_3=E_{22}X_2-X_2E_{11}-X_1E_{12}X_1\ ,$$

and, in general, for $k=3,4,\dots$,

$$\begin{aligned} X_k:\quad X_kM_{11}-M_{22}X_k=&E_{22}X_{k-1}-X_{k-1}E_{11}+\\ &-\left[X_{k-2}E_{12}X_1+X_{k-3}E_{12}X_2+\cdots+X_1E_{12}X_{k-2}\right]\ .\end{aligned}\tag{A.4}$$

Therefore, because of (A.2), the terms X_k in the expansion (A.3) are well defined and are $\mathcal{C}^k$ functions for $(\alpha,\beta,\gamma)\in B$. Observe that each of the terms X_k satisfies the same linear system with different right-hand sides. That is, we can write $X_k=\mathcal{S}^{-1}C_k$, where $\mathcal{S}^{-1}$ is the inverse of the Sylvester operator $Z\to ZM_{11}-M_{22}Z$ and C_k express the right-hand sides of the recursion (A.4).

Let $\sigma=\max_B\|\mathcal{S}^{-1}\|$, so that we can write $\|X_k\|\le\sigma\|C_k\|$. Our next goal is to show that

$$\|C_k\|\le c_k\sigma^{k-1}\eta^k\,,\qquad\text{where}\qquad 0<c_k\le 4^k\,,\ \ k=1,2,\dots\ .\tag{A.5}$$

As soon as (A.5) is verified, we can complete the proof as follows.

(i) From (A.5), the series (A.3) converges uniformly as long as $4\varepsilon\sigma\eta<1$, a condition that can be always satisfied in a sufficiently small ε-neighborhood of 0.

(ii) From uniform convergence of (A.3), we will have that the limit function is continuous in $(\alpha,\beta,\gamma)\in B$.

(iii) Since $\widetilde{X}$ of point **1.** above is the unique continuous solution of (A.1) passing through $X=0$ for $\varepsilon=0$, in sufficiently small neighborhoods of the origin, then we must have that $\widetilde{X}$ is the same as X given by (A.3), in sufficiently small neighborhoods of the origin. And, therefore, in these neighborhoods, the function $\widetilde{X}$ is given by (A.3) and it identifies the unique solution of (A.1), analytic in ε (because of (A.3)) and $\mathcal{C}^k$ in (α,β,γ), which is what we wanted to prove.

Finally, let us verify (A.5). Taking norms, from (A.4) inductively we get

$$
\begin{aligned}
&\|X_k\| \le \sigma \|C_k\| \ , \quad \text{and} \\
&\|C_k\| \le \eta \Big(2\|X_{k-1}\| + \|X_{k-2}\|\|X_1\| + \dots + \|X_1\|\|X_{k-2}\| \Big) \\
&\qquad \le \sigma\eta \Big(2\|C_{k-1}\| + \|C_{k-2}\|\|C_1\| + \|C_{k-3}\|\|C_2\| + \dots + \|C_1\|\|C_{k-2}\| \Big) \\
&\qquad \le \sigma^{k-1}\eta^k \Big(2c_{k-1} + c_{k-2}c_1 + c_{k-3}c_2 + \dots + c_1 c_{k-2} \Big) \ ,
\end{aligned}
$$

and so the issue has become to study the growth of the numerical sequence recursively defined by

$$
c_k = c_{k-1}c_0 + c_{k-2}c_1 + \dots + c_1 c_{k-2} + c_0 c_{k-1} = \sum_{j=0}^{k-1} c_{k-1-j}c_j \ , \ \ k = 2,3,4,\dots, \quad c_0 = c_1 = 1 \, .
$$

Now, suppose that we had a function $f(u)$ with convergent McLaurin's expansion $f(u) = \sum_{k=0}^{\infty} c_k u^k$. Then, the expansion for $f^2(u)$ would be $(c_0+c_1u+c_2u^2+\dots)(c_0+c_1u+c_2u^2+\dots) = c_0^2 + (c_1c_0+c_0c_1)u + \dots + \sum_{j=0}^{k-1} c_{k-1-j}c_j u^k + \dots$, from which we then recognize that the coefficients c_k's we are after are simply the coefficients in the power series expansion of the function $f(u) = \frac{1-\sqrt{1-4u}}{2u}$, which has radius of convergence $1/4$. By using the power series expansion for the square root, and simplifying, we obtain

$$
c_k = \frac{1}{2}(-1)^k \binom{1/2}{k+1} 4^{k+1} \ , \qquad \text{where} \qquad \binom{1/2}{j} = \frac{1/2(1/2-1)\dots(1/2-j+1)}{j!} \ ,
$$

and therefore $c_{k+1}/c_k = 4 - \frac{6}{k+2}$ and (A.5) is proved. □

Remark A.2. An interesting aspect of Theorem A.1 is that it validates an implicit function theorem giving analyticity with respect to a parameter (ε), and $\mathcal{C}^k$ smoothness with respect to the other parameters.

Proof of Theorem 2.10. Finally, let us give a complete proof of Theorem 2.10.

We now proceed as follows.

(1) We seek the transformation $T = \begin{bmatrix} I & 0 \\ X & I \end{bmatrix}$, with X taking values in $\mathbb{R}^{2n-4,4}$, such that

$$
\begin{aligned}
T^{-1} \begin{bmatrix} M_1 + \varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2 + \varepsilon H \end{bmatrix} T &= \begin{bmatrix} I & 0 \\ -X & I \end{bmatrix} \begin{bmatrix} M_1 + \varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2 + \varepsilon H \end{bmatrix} \begin{bmatrix} I & 0 \\ X & I \end{bmatrix} = \\
&= \begin{bmatrix} M_1 + \varepsilon\widehat{E} + \varepsilon F X & \varepsilon F \\ 0 & M_2 + \varepsilon H - \varepsilon X F \end{bmatrix} .
\end{aligned} \tag{A.6}
$$

Observe that this transformation T exists if and only X satisfies the following special version of (A.1):

$$
\varepsilon F^T - X(M_1 + \varepsilon\widehat{E}) + (M_2 + \varepsilon H)X - \varepsilon X F X \ = \ 0 \, . \tag{A.7}
$$

Thus, because of Theorem A.1, the transformation is well defined (in appropriate neighborhoods of the origin) and X has the form as in (A.3).

(2) Next, we consider the following block-QR factorization (cfr. [19]):

$$\begin{aligned} &\begin{bmatrix} I & 0 \\ X & I \end{bmatrix} = QR\,, \quad \text{where} \quad Q = \begin{bmatrix} I & -X^T \\ X & I \end{bmatrix} \begin{bmatrix} (I+X^TX)^{-1/2} & 0 \\ 0 & (I+XX^T)^{-1/2} \end{bmatrix} \\ &\qquad\qquad \text{and} \qquad R = \begin{bmatrix} (I+X^TX)^{1/2} & (I+X^TX)^{-1/2}X^T \\ 0 & (I+XX^T)^{-1/2} \end{bmatrix}\,, \end{aligned} \tag{A.8}$$

where we have taken the unique positive definite square root; note that $I + X^TX$ and $I+XX^T$ are obviously positive definite and the unique positive definite square root of these functions is as smooth as the functions themselves (see also (A.15)). Now, observe that Q is orthogonal ($Q^TQ = I_{2n}$) and therefore we must have that

$$Q^T \begin{bmatrix} M_1 + \varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2 + \varepsilon H \end{bmatrix} Q$$

is both symmetric and (block) upper triangular. Then, it must be block diagonal. That is, we have

$$\begin{aligned} &Q^T \begin{bmatrix} M_1 + \varepsilon\widehat{E} & \varepsilon F \\ \varepsilon F^T & M_2 + \varepsilon H \end{bmatrix} Q = \begin{bmatrix} \widetilde{M_1} & 0 \\ 0 & \widetilde{M_2} \end{bmatrix}\,, \quad \text{where } \widetilde{M_1} \text{ is given by} \\ &(I+X^TX)^{-1/2}\Big[M_1 + \varepsilon\widehat{E} + \varepsilon(X^TF^T + FX) + X^T(M_2+\varepsilon H)X\Big](I+X^TX)^{-1/2} \end{aligned} \tag{A.9}$$

and a similar expression for $\widetilde{M_2}$. We focus on $\widetilde{M_1}$ only, since we are only interested in tracking the eigenvalues of $\widetilde{M_1}$, but the argument for $\widetilde{M_2}$ is virtually identical.

(3) Next, we take a closer look at the form of the solution X of (A.7). Because of Theorem A.1, we know that for this X we have the expression

$$\begin{aligned} &X = \sum_{k=1}^{\infty} \varepsilon^k X_k, \qquad \text{where} \\ &X_1: \quad X_1M_1 - M_2X_1 = F^T\,, \quad \text{and} \\ &X_k: \quad X_kM_1 - M_2X_k = R_k\,, \quad \text{for } k = 2,3,\dots, \quad \text{where} \\ &\qquad R_k = HX_{k-1} - X_{k-1}\widehat{E} - [X_{k-2}FX_1 + X_{k-3}FX_2 + \cdots + X_1FX_{k-2}]\,. \end{aligned} \tag{A.10}$$

The following construction clarifies the form of X.

Definition A.3. Consider a matrix $V \in \mathbb{R}^{2n-4,4}$, $n \geq 3$, and partition it as $V = \begin{bmatrix} V_1 & V_2 \\ V_3 & V_4 \end{bmatrix}$, where $V_1, V_2, V_3, V_4 \in \mathbb{R}^{n-2,2}$. We say that V is of "type-E" if $V = \begin{bmatrix} V_1 & V_2 \\ V_2 & -V_1 \end{bmatrix}$, and we say that V is of "type-M" if $V = \begin{bmatrix} V_1 & V_2 \\ -V_2 & V_1 \end{bmatrix}$.

Clearly, type-E and type-M matrices form subspaces of $\mathbb{R}^{2n-4,4}$, which we will call $\mathcal{S}_E$ and $\mathcal{S}_M$ respectively. Moreover, they are mutually complementary, since we can consider the unique decomposition (projection) of any matrix $V \in \mathbb{R}^{2n-4,4}$ into the sum of a type-E and a type-M matrix, simply taking $V = Y + Z$ with $Y = \begin{bmatrix} (V_1-V_4)/2 & (V_2+V_3)/2 \\ (V_2+V_3)/2 & -(V_1-V_4)/2 \end{bmatrix}$ and $Z = \begin{bmatrix} (V_1+V_4)/2 & (V_2-V_3)/2 \\ -(V_2-V_3)/2 & (V_1+V_4)/2 \end{bmatrix}$. Observe that F^T in (A.6) and (A.7) is of type-E, i.e. $F^T \in \mathcal{S}_E$.

In particular, for the function X solution of (A.7), we can write

$$X = Y + Z = \sum_{k=1}^{\infty} \varepsilon^k (Y_k + Z_k)\,, \qquad \text{where} \qquad Y_k \in \mathcal{S}_E\ ,\ Z_k \in \mathcal{S}_M\ ,\ \ k = 1, 2, \dots\ . \tag{A.11}$$

The following two Lemmata can be proved by (tedious, but straightforward, computations) directly multiplying the matrices involved.

Lemma A.4. *Let M_1, M_2, E, H be the matrices in* (A.6)*.*

The Sylvester operator $V \to VM_1 - M_2V$ leave $\mathcal{S}_E$ and $\mathcal{S}_M$ invariant. In other words, if $Y \in \mathcal{S}_E$, then $YM_1 - M_2Y \in \mathcal{S}_E$, if $Z \in \mathcal{S}_M$, then $ZM_1 - M_2Z \in \mathcal{S}_M$.

Similarly, the Sylvester operator $V \to VE - HV$ transforms $\mathcal{S}_E$ into $\mathcal{S}_M$, and viceversa. In other words, if $Y \in \mathcal{S}_E$, then $YE - HY \in \mathcal{S}_M$, if $Z \in \mathcal{S}_M$, then $ZE - HZ \in \mathcal{S}_E$. □

Lemma A.5. *Let $Y, \widehat{Y} \in \mathcal{S}_E$ and $Z, \widehat{Z} \in \mathcal{S}_M$, and let $F^T \in \mathcal{S}_E$. Then:*
(a) *$ZFY + YFZ \in \mathcal{S}_M$;*
(b) *$ZFZ \in \mathcal{S}_E$ and $YFY \in \mathcal{S}_E$;*
(c) *$ZF\widehat{Z} + \widehat{Z}FZ \in \mathcal{S}_E$, $YF\widehat{Y} + \widehat{Y}FY \in \mathcal{S}_E$.* □

With these, we can now prove the following structural result on the solution X of (A.7).

Lemma A.6. *Let X be given by* (A.10) *and further rewritten as in* (A.11)*. Then: $Y_k = 0$, for k even, and $Z_k = 0$ for k odd. In other words, the solution of* (A.10) *can be written as:*

$$\begin{gathered} X = Y + Z\ ,\ Y = \sum_{k=1}^{\infty} \varepsilon^{2k-1} Y_{2k-1}\ ,\ Z = \sum_{k=1}^{\infty} \varepsilon^{2k} Z_{2k}\,, \\ \textit{where}\ \ Y_{2k-1} \in \mathcal{S}_E\ ,\ Z_{2k} \in \mathcal{S}_M\,,\quad \textit{for all}\ k = 1, \dots,\ . \end{gathered} \tag{A.12}$$

Proof. The proof is by induction on the index k in the summation of (A.11).

For $k = 1$, we simply have $X_1M_1 - M_2X_1 = F^T$. Since $F^T \in \mathcal{S}_E$, because of Lemma A.4 we get $X_1 = Y_1$.

Next, assuming the result to be true up to index k, for X_k we have $X_kM_1 - M_2X_k = R_k$. But, by looking at the form of R_k, using the induction hypothesis, Lemma A.4 and (repeatedly) Lemma A.5, we can conclude that $R_k \in \mathcal{S}_E$ for k odd, and $R_k \in \mathcal{S}_M$ for k even. Therefore, using Lemma A.4 we get that $Y_k = 0$ for k odd and $Z_k = 0$ for k even, as claimed. □

(4) Finally, we look at the term $\widetilde{M_1}$ in (A.9). First, consider the term $(I+X^TX)^{1/2}\widetilde{M_1}(I+X^TX)^{1/2}$, that is

$$M_1 + \varepsilon\widehat{E} + \varepsilon(X^TF^T + FX) + X^T(M_2 + \varepsilon H)X. \tag{A.13}$$

Definition A.7. A matrix $V \in \mathbb{R}^{4\times 4}$, symmetric ($V^T = V$), is called "E-like" if $V = \begin{bmatrix} V_1 & V_2 \\ V_2 & -V_1 \end{bmatrix}$, $V_i^T = V_i$, $i = 1, 2$, and "M-like" if $V = \begin{bmatrix} V_1 & V_2 \\ -V_2 & V_1 \end{bmatrix}$, $V_1^T = V_1$ and $V_2^T = -V_2$; here, $V_i \in \mathbb{R}^{2\times 2}$, $i = 1, 2$.

Observe that M_1 is "M-like", and E is "E-like", and that there is a unique decomposition of symmetric (4×4) matrices into the sum of "M-like" and "E-like" matrices (they form subspaces, which are obviously invariant under transposition).

The following Lemma is useful to complete our argument, and it can be easily proved by direct verification.

Lemma A.8. *Let $M_1, M_2, E, H, F,$ be the matrices in* (A.13), *with X given by* (A.12). *Then:*

(a) *$(FY + Y^TF^T)$ is "M-like", and $(ZY + Z^TF^T)$ is "E-like";*

(b) *Y^TM_2Y and Z^TM_2Z are "M-like", whereas Z^TM_2Y and Y^TM_2Z are "E-like";*

(c) *Y^THY and Z^THZ are "E-like", whereas Z^THY and Y^THZ are "M-like";*

(d) *Let P be positive definite and write $P = P_M + P_E$, where P_M is "M-like" and P_E is "E-like". Further, let V be "M-like" and W be "E-like". Then:*

(d-1) *P_MVP_M is "M-like" and P_EVP_E is "E-like";*

(d-2) *P_MWP_M is "E-like" and P_EWP_E is "M-like";*

(d-3) *$P_MVP_E+P_EVP_M$ is "E-like" and $P_MWP_E+P_EWP_M$ is "M-like".* □

Using Lemma A.8, and the form of X from (A.12), we immediately obtain that the term in (A.13) can be written as

$$(M_1 + \sum_{k=1}^{\infty} \varepsilon^{2k}\widehat{\widehat{E}}_{2k}) + \varepsilon(\widehat{E} + \sum_{k=1}^{\infty} \varepsilon^{2k}\widehat{\widehat{E}}_{2k+1}) \ ,$$

where, for $k = 1, 2, \dots$, each $\widehat{\widehat{E}}_{2k}$ is "M-like" and each $\widehat{\widehat{E}}_{2k+1}$ is "E-like".

Next, we notice that no structural change takes place when forming $\widetilde{M_1}$:

$$(I + X^TX)^{-1/2}\left[(M_1 + \sum_{k=1}^{\infty} \varepsilon^{2k}\widehat{\widehat{E}}_{2k}) + \varepsilon(\widehat{E} + \sum_{k=1}^{\infty} \varepsilon^{2k}\widehat{\widehat{E}}_{2k+1})\right](I + X^TX)^{-1/2} \ ; \tag{A.14}$$

this is because $(I+X^TX)^{-1/2}$ has the following series expansion (norm-convergent for $\|X^TX\| < 1$, which can be trivially guaranteed for ε sufficiently small because of (A.11)):

$$(I + X^TX)^{-1/2} \ = \ \sum_{k=0}^{\infty} \binom{-1/2}{k}(X^TX)^k \, . \tag{A.15}$$

With this, writing $X = Y + Z$ as in (A.12), and observing that expressions of the type $Y_j^T Z_l + Z_l^T Y_j$ are "E-like", whereas terms of the type $Y_j^T Y_j$ and $Z_l^T Z_l$ are "M-like", repeatedly using Lemma A.8, we have completed the proof of Theorem 2.10. □

References

[1] M. Baer, *Beyond Born-Oppenheimer: electronic nonadiabatic coupling terms and conical intersections*. John Wiley & Sons, Hoboken, NJ, 2006.

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

[3] M. V. Berry, Physics of nonhermitian degeneracies. *Czechoslovak Journal of Physics*, 54–10: 1039–1047, 2004.

[4] L. Dieci and A. Papini and A. Pugliese, Approximating coalescing points for eigenvalues of Hermitian matrices of three parameters. *SIAM J. Matrix Anal. Appl.*, 34(2):519–541, 2013.

[5] L. Dieci and A. Pugliese, Two-parameter SVD: Coalescing singular values and periodicity. *SIAM J. Matrix Anal. Appl.*, 31:375–403, 2009.

[6] L. Dieci and A. Pugliese, Singular values of two-parameter matrices: an algorithm to accurately find their intersections. *Math. Comput. Simulation*, 79(4):1255–1269, 2008.

[7] L. Dieci and A. Pugliese, Hermitian matrices depending on three parameters: Coalescing eigenvalues. *Linear Algebra and Its Applications*, 436; 4120–4142, 2012.

[8] A. Gallina, L. Pichler, and T. Uhl, Enhanced meta-modelling technique for analysis of mode crossing, mode veering and mode coalescence in structural dynamics. *Mechanical Systems and Signal Processing*, 25:2297–2312, 2011.

[9] H. Gingold, A method of global blockdiagonalization for matrix-valued functions. *SIAM J. Math. Anal.*, 9:1076–1082, 1978.

[10] M. W. Hirsch, *Differential Topology*, Springer-Verlag, New–York, 1976.

[11] P. F. Hsieh and Y. Sibuya, *A global analysis of matrices of functions of several variables*, J. Math. Anal. Appl., 14 (1966), pp. 332–340.

[12] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1976. 2nd edition.

[13] H. B. Keller, *Lectures on numerical methods in bifurcation problems*, volume 79 of *Tata Institute of Fundamental Research Lectures on Mathematics and Physics*. Published for the Tata Institute of Fundamental Research, Bombay, 1987.

[14] O.N. Kirillov, A.A. Mailybaev, and A.P. Seyranian, Unfolding of eigenvalue surfaces near a diabolic point due to a complex perturbation. *J. Phys. A: Math. Gen.*, 38:5531–5546, 2005.

[15] A.A. Mailybaev, O.N. Kirillov, and A.P. Seyranian, Strong and weak coupling of eigenvalues of complex matrices. *Proceedings of Physics and Control, International Conference*, 312–318, 2005.

[16] A.A. Mailybaev, O.N. Kirillov, and A.P. Seyranian, Berry phase around degeneracies. *Doklady Mathematics*, 73-1:129–133, 2006.

[17] R. Melville and D. S. Mackey, A new algorithm for two-dimensional numerical continuation. *Comput. Math. Appl.*, 30(1):31–46, 1995.

[18] A. Srikantha Phani, J. Woodhouse, and N.A. Fleck, Wave propagation in two-dimensional periodic lattices. *J. Acoust. Soc. Am.*, 119:1995–2005, 2006.

[19] G. W. Stewart, Error and perturbation bounds for subspaces associated with certain eigenvalue problems. *SIAM Rev.*, 15:727–764, 1973.

[20] G. W. Stewart and J. G. Sun, *Matrix Perturbation Theory.* Academic Press, San Diego, 1990.

[21] A. J. Stone, Spin-Orbit Coupling and the Intersection of Potential Energy Surfaces in Polyatomic Molecules. *Proc. Roy. Soc. Lond.*, A351:141–150, 1976.

[22] J. von Neumann and E. Wigner, Eigenwerte bei adiabatischen prozessen. *Physik Zeitschrift*, 30:467–470, 1929.

[23] P.N. Walker, M.J. Sanchez, and M. Wilkinson, Singularities in the spectra of random matrices. *J. Mathem. Phys.*, 37-10:5019–5032, 1996.

[24] David R. Yarkony, Conical intersections: The new conventional wisdom. *J. Phys. Chem. A*, 105:6277–6293, 2001.

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

Dipartimento di Matematica, Università degli Studi di Bari Aldo Moro, Via Orabona 4, Bari, 70125 Italy
E-mail address: alessandro.pugliese@uniba.it