International Journal of Applied Mathematical Sciences.
ISSN 0973-0176 Vol.2 No.2 (2005), pp. 159-171

http://www.gbspublisher.com/ijams.htm

Pseudoinverse Formulation of Rayleigh-Schrödinger Perturbation Theory for the Symmetric Definite Generalized Eigenvalue Problem

Brian J. McCartin
Applied Mathematics
Kettering University
1700 West Third Avenue
Flint, MI 48504-4898 USA
bmccarti@kettering.edu

Abstract

In this paper, a comprehensive treatment of Rayleigh-Schrödinger perturbation theory for the symmetric definite generalized eigenvalue problem [1, 2] is furnished with emphasis on the degenerate problem. The treatment is simply based upon the Moore-Penrose pseudoinverse thus constituting the natural generalization of the procedure for the standard symmetric eigenvalue problem [3]. In addition to providing a concise matrix-theoretic formulation of this procedure, it also provides for the explicit determination of that stage of the algorithm where each higher order eigenvector correction becomes fully determined. Along the way, we generalize the Dalgarno-Stewart identities [15] from the standard to the generalized eigenvalue problem. The general procedure is illustrated by an extensive example.
2000 Mathematics Subject Classification: 15A18, 65F15.

Keywords: spectral perturbation theory, symmetric definite generalized eigenvalue problem, Moore-Penrose pseudoinverse.

1. Introduction

A comprehensive treatment of Rayleigh-Schrödinger [1, 2] perturbation theory for the symmetric matrix eigenvalue problem based upon the Moore-Penrose pseudoinverse was provided in [3]. It is the express intent of the present paper to extend this technique to the symmetric definite generalized eigenvalue problem. The origin of such problems in the analysis of electromechanical systems is discussed in [4].

Mathematically, we have a discretized differential operator embodied in a real symmetric matrix pair, (A_0, B_0) with B_0 positive definite, which is subjected to a small symmetric linear perturbation, $(A, B) = (A_0 + \varepsilon A_1, B_0 + \varepsilon B_1)$ with B also positive definite, due to physical inhomogeneities. The Rayleigh-Schrödinger procedure produces approximations to the eigenvalues and eigenvectors of (A, B) by a sequence of successively higher order corrections to the eigenvalues and eigenvectors of (A_0, B_0). Observe that $B_1 = 0$ permits reduction to the standard eigenvalue problem $(B^{-1}A_0 + \varepsilon B^{-1}A_1, I)$. However, this destroys the very symmetry which is the linchpin of the Rayleigh-Schrödinger procedure.

The difficulty with standard treatments of this procedure [5] is that the eigenvector corrections are expressed in a form requiring the complete collection of eigenvectors of (A_0, B_0). For large matrices this is clearly an undesirable state of affairs. Consideration of the thorny issue of multiple eigenvalues of (A_0, B_0) [6] only serves to exacerbate this difficulty.

This malady can be remedied by expressing the Rayleigh-Schrödinger procedure in terms of the Moore-Penrose pseudoinverse [7]. This permits these corrections to be computed knowing only the eigenvectors of (A_0, B_0) corresponding to the eigenvalues of interest. In point of fact, the pseudoinverse need not be explicitly calculated since only pseudoinverse-vector products are required. In turn, these may be efficiently calculated by a combination of LU-factorization and orthogonal projections. However, the formalism of the pseudoinverse provides a concise formulation of the procedure and permits ready analysis of theoretical properties of the algorithm.

Since the present paper is only concerned with the real symmetric definite case, the existence of a complete set of B-orthonormal eigenvectors is assured [8, 9, 10]. The much more difficult case of defective matrices has been considered in [11] for the standard eigenvalue problem. Moreover, we only consider the computational aspects of this procedure. Existence of the relevant perturbation expansions follows from the rigorous theory developed in [12, 13, 14].

2. Nondegenerate Case

Consider the generalized eigenvalue problem

$$Ax_i = \lambda_i B x_i \ (i = 1, \ldots, n), \tag{1}$$

where A and B are real, symmetric, $n \times n$ matrices and B is further assumed to be positive definite. We also assume that this matrix pair has distinct eigenvalues, λ_i $(i = 1, \ldots, n)$. Under these assumptions the eigenvalues are real and the corresponding eigenvectors, x_i $(i = 1, \ldots, n)$, are guaranteed to be B-orthogonal [4, 7, 10].

Next, let

$$A(\varepsilon) = A_0 + \varepsilon A_1; \ B(\varepsilon) = B_0 + \varepsilon B_1, \tag{2}$$

where, likewise, A_0 is real and symmetric and B_0 is real, symmetric and positive definite, except that now the matrix pair, (A_0, B_0), may possess multiple eigenvalues (called degeneracies in the physics literature). The root cause of such degeneracy is typically the presence of some underlying symmetry. Any attempt to weaken the assumption on the eigenstructure of (A,B) leads to a Rayleigh-Schrödinger iteration that never terminates [13, p. 92]. In the remainder of this section, we consider the nondegenerate case where the unperturbed eigenvalues, $\lambda_i^{(0)}$ $(i=1,\ldots,n)$, are all distinct. Consideration of the degenerate case is deferred to the next section.

Under the above assumptions, it is shown in [12, 13, 14] that the eigenvalues and eigenvectors of (A,B) possess the respective perturbation expansions,

$$\lambda_i(\varepsilon)=\sum_{k=0}^{\infty}\varepsilon^k\lambda_i^{(k)};\ x_i(\varepsilon)=\sum_{k=0}^{\infty}\varepsilon^k x_i^{(k)}\ (i=1,\ldots,n), \tag{3}$$

for sufficiently small ε and $B=I$. Using the Cholesky factorization, $B=LL^T$, this theory may be straightforwardly extended to accomodate arbitrary symmetric positive definite B [4]. Importantly, it is not necessary to actually calculate the Cholesky factorization of B in the computational procedure developed below. Clearly, the zeroth-order terms, $\{\lambda_i^{(0)};\ x_i^{(0)}\}$ $(i=1,\ldots,n)$, are the eigenpairs of the unperturbed matrix pair, (A_0,B_0). I.e.,

$$(A_0-\lambda_i^{(0)}B_0)x_i^{(0)}=0\ (i=1,\ldots,n). \tag{4}$$

The unperturbed mutually B_0-orthogonal eigenvectors, $x_i^{(0)}$ $(i=1,\ldots,n)$, are assumed to have been B_0-normalized to unity.

Substitution of Equations (2) and (3) into Equation (1) yields the recurrence relation

$$(A_0-\lambda_i^{(0)}B_0)x_i^{(k)}=-(A_1-\lambda_i^{(1)}B_0-\lambda_i^{(0)}B_1)x_i^{(k-1)}+\sum_{j=0}^{k-2}(\lambda_i^{(k-j)}B_0+\lambda_i^{(k-j-1)}B_1)x_i^{(j)}, \tag{5}$$

for $(k=1,\ldots,\infty;\ i=1,\ldots,n)$. For fixed i, solvability of Equation (5) requires that its right hand side be orthogonal to $x_i^{(0)}$ for all k. Thus, the value of $x_i^{(j)}$ determines $\lambda_i^{(j+1)}$. Specifically,

$$\lambda_i^{(j+1)}=\langle x_i^{(0)},A_1x_i^{(j)}\rangle-\sum_{l=0}^{j}\lambda_i^{(j-l)}\langle x_i^{(0)},B_1x_i^{(l)}\rangle, \tag{6}$$

where we have employed the so-called **intermediate normalization** that $x_i^{(k)}$ shall be chosen to be B_0-orthogonal to $x_i^{(0)}$ for $k=1,\ldots,\infty$. This is equivalent to $\langle x_i^{(0)},B_0x_i(\varepsilon)\rangle=1$ and this normalization will be used throughout the remainder of this work.

For the standard eigenvalue problem, $B=I$, a beautiful result due to Dalgarno and Stewart [15] (sometimes incorrectly attributed to Wigner in the physics literature) says that much more is true: The value of the eigenvector correction $x_i^{(j)}$, in fact, determines the eigenvalues through $\lambda_i^{(2j+1)}$. For the generalized eigenvalue problem, Equation (1), this may be generalized by the following constructive procedure which heavily exploits the symmetry of A_0, A_1, B_0, and B_1.

We commence by observing that

$$\begin{aligned}\lambda_i^{(k)} &= \langle x_i^{(0)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(k-1)} \rangle - \textstyle\sum_{l=1}^{k-1} \lambda_i^{(l)} \langle x_i^{(0)}, B_1 x_i^{(k-l-1)} \rangle \\ &= \langle x_i^{(k-1)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(0)} \rangle - \textstyle\sum_{l=1}^{k-1} \lambda_i^{(l)} \langle x_i^{(0)}, B_1 x_i^{(k-l-1)} \rangle \\ &= -\langle x_i^{(k-1)}, (A_0 - \lambda_i^{(0)} B_0) x_i^{(1)} \rangle - \textstyle\sum_{l=1}^{k-1} \lambda_i^{(l)} \langle x_i^{(0)}, B_1 x_i^{(k-l-1)} \rangle \\ &= -\langle x_i^{(1)}, (A_0 - \lambda_i^{(0)} B_0) x_i^{(k-1)} \rangle - \textstyle\sum_{l=1}^{k-1} \lambda_i^{(l)} \langle x_i^{(0)}, B_1 x_i^{(k-l-1)} \rangle \\ &= \langle x_i^{(1)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(k-2)} \rangle - \lambda_i^{(1)} \langle x_i^{(0)}, B_1 x_i^{(k-2)} \rangle \\ &- \textstyle\sum_{l=2}^{k-1} [\langle x_i^{(1)}, (\lambda_i^{(l)} B_0 + \lambda_i^{(l-1)} B_1) x_i^{(k-l-1)} \rangle + \lambda_i^{(l)} \langle x_i^{(0)}, B_1 x_i^{(k-l-1)} \rangle].\end{aligned} \tag{7}$$

Continuing in this fashion, we eventually arrive at, for even $k = 2j$ $(j = 2, \ldots)$,

$$\begin{aligned}\lambda_i^{(2j)} &= \langle x_i^{(j-1)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(j)} \rangle \\ &- \textstyle\sum_{\mu=2}^{j-1} \lambda_i^{(\mu)} \sum_{\nu=j-\mu+1}^{j} \langle x_i^{(2j-\mu-\nu)}, B_0 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=j}^{2j-2} \lambda_i^{(\mu)} \sum_{\nu=1}^{2j-\mu-1} \langle x_i^{(2j-\mu-\nu)}, B_0 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=1}^{j-1} \lambda_i^{(\mu)} \sum_{\nu=j-\mu}^{j} \langle x_i^{(2j-\mu-\nu-1)}, B_1 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=j}^{2j-1} \lambda_i^{(\mu)} \sum_{\nu=0}^{2j-\mu-1} \langle x_i^{(2j-\mu-\nu)}, B_1 x_i^{(\nu)} \rangle,\end{aligned} \tag{8}$$

while, for odd $k = 2j + 1$ $(j = 1, \ldots)$,

$$\begin{aligned}\lambda_i^{(2j+1)} &= \langle x_i^{(j)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(j)} \rangle \\ &- \textstyle\sum_{\mu=2}^{j-1} \lambda_i^{(\mu)} \sum_{\nu=j-\mu+1}^{j} \langle x_i^{(2j-\mu-\nu+1)}, B_0 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=j}^{2j-1} \lambda_i^{(\mu)} \sum_{\nu=1}^{2j-\mu} \langle x_i^{(2j-\mu-\nu+1)}, B_0 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=1}^{j-1} \lambda_i^{(\mu)} \sum_{\nu=j-\mu}^{j} \langle x_i^{(2j-\mu-\nu)}, B_1 x_i^{(\nu)} \rangle \\ &- \textstyle\sum_{\mu=j}^{2j} \lambda_i^{(\mu)} \sum_{\nu=0}^{2j-\mu} \langle x_i^{(2j-\mu-\nu)}, B_1 x_i^{(\nu)} \rangle,\end{aligned} \tag{9}$$

where the second sum is to be omitted for $j = 1$. This important pair of equations will henceforth be referred to as the **generalized Dalgarno-Stewart identities**. $\lambda_i^{(1)}$ and $\lambda_i^{(2)}$ may be obtained directly from Equation (6).

The eigenvector corrections are determined recursively from Equation (5) as

$$x_i^{(k)} = (A_0 - \lambda_i^{(0)} B_0)^\dagger [-(A_1 - \lambda_i^{(1)} B_0 - \lambda_i^{(0)} B_1) x_i^{(k-1)} + \sum_{j=0}^{k-2} (\lambda_i^{(k-j)} B_0 + \lambda_i^{(k-j-1)} B_1) x_i^{(j)}], \tag{10}$$

for $(k = 1, \ldots, \infty;\ i = 1, \ldots, n)$, where $(A_0 - \lambda_i^{(0)} B_0)^\dagger$ denotes the Moore-Penrose pseudoinverse [7] of $(A_0 - \lambda_i^{(0)} B_0)$ and intermediate normalization has been employed.

3. Degenerate Case

When the matrix pair (A_0, B_0) possesses multiple eigenvalues (the so-called degenerate case), the above straightforward analysis for the nondegenerate case encounters serious complications. This is a consequence of the fact that, in this new case, Rellich's Theorem [12, pp. 42-45] guarantees the existence of the perturbation expansions, Equation (3), only for certain special unperturbed eigenvectors. These special unperturbed eigenvectors cannot be specified *a priori* but must instead emerge from the perturbation procedure itself.

Furthermore, the higher order corrections to these special unperturbed eigenvectors are more stringently constrained than previously since they must be chosen so that Equation (5) is always solvable. I.e., they must be chosen so that the right hand side of Equation (5) is always orthogonal to the entire eigenspace associated with the multiple eigenvalue in question.

Thus, without any loss of generality, suppose that $\lambda_1^{(0)} = \lambda_2^{(0)} = \cdots = \lambda_m^{(0)} = \lambda^{(0)}$ is just such an eigenvalue of multiplicity m with corresponding known B_0-orthonormal eigenvectors $x_1^{(0)}, x_2^{(0)}, \ldots, x_m^{(0)}$. Then, we are required to determine appropriate linear combinations

$$y_i^{(0)} = a_1^{(i)} x_1^{(0)} + a_2^{(i)} x_2^{(0)} + \cdots + a_m^{(i)} x_m^{(0)} \ (i = 1, \ldots, m) \tag{11}$$

so that the expansions, Equation (3), are valid with $x_i^{(k)}$ replaced by $y_i^{(k)}$. **In point of fact, the remainder of this paper will assume that x_i has been replaced by y_i in Equations (3)-(10).** Moreover, the higher order eigenvector corrections, $y_i^{(k)}$, must be suitably determined. Since we would like $\{y_i^{(0)}\}_{i=1}^m$ to be likewise B_0-orthonormal, we require that

$$a_1^{(\mu)} a_1^{(\nu)} + a_2^{(\mu)} a_2^{(\nu)} + \cdots + a_m^{(\mu)} a_m^{(\nu)} = \delta_{\mu,\nu} \ . \tag{12}$$

Recall that we have assumed throughout that the perturbed matrix pair, $(A(\varepsilon), B(\varepsilon))$, itself has distinct eigenvalues, so that eventually all such degeneracies will be fully resolved. What significantly complicates matters is that it is not known beforehand at what stages portions of the degeneracy will be resolved.

In order to bring order to a potentially calamitous situation, we will begin by first considering the case where the degeneracy is fully resolved at first order. Only then do we move on to study the case where the degeneracy is completely and simultaneously resolved at Nth order. Finally, we will have laid sufficient groundwork to permit treatment of the most general case of mixed degeneracy where resolution occurs across several different orders. This seems preferable to presenting an impenetrable collection of opaque formulae.

3.1. First Order Degeneracy

We first dispense with the case of first order degeneracy wherein $\lambda_i^{(1)}$ $(i = 1\ldots,m)$ are all distinct. In this event, we determine $\{\lambda_i^{(1)}; y_i^{(0)}\}_{i=1}^m$ by insisting that Equation (5) be solvable for $k = 1$ and $i = 1,\ldots,m$. In order for this to obtain, it is both necessary and sufficient that, for each fixed i,

$$\langle x_\mu^{(0)}, (A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(0)} \rangle = 0 \; (\mu = 1,\ldots,m). \tag{13}$$

Inserting Equation (11) and invoking the B_0-orthonormality of $\{x_\mu^{(0)}\}_{\mu=1}^m$, we arrive at, in matrix form,

$$\begin{bmatrix} \langle x_1^{(0)}, (A_1 - \lambda^{(0)} B_1) x_1^{(0)} \rangle & \cdots & \langle x_1^{(0)}, (A_1 - \lambda^{(0)} B_1) x_m^{(0)} \rangle \\ \vdots & \ddots & \vdots \\ \langle x_m^{(0)}, (A_1 - \lambda^{(0)} B_1) x_1^{(0)} \rangle & \cdots & \langle x_m^{(0)}, (A_1 - \lambda^{(0)} B_1) x_m^{(0)} \rangle \end{bmatrix} \begin{bmatrix} a_1^{(i)} \\ \vdots \\ a_m^{(i)} \end{bmatrix} = \lambda_i^{(1)} \begin{bmatrix} a_1^{(i)} \\ \vdots \\ a_m^{(i)} \end{bmatrix}. \tag{14}$$

Thus, each $\lambda_i^{(1)}$ is an eigenvalue with corresponding eigenvector $[a_1^{(i)},\ldots,a_m^{(i)}]^T$ of the matrix M defined by $M_{\mu,\nu} = \langle x_\mu^{(0)}, M^{(1)} x_\nu^{(0)} \rangle$ $(\mu,\nu = 1,\ldots,m)$ where $M^{(1)} := A_1 - \lambda^{(0)} B_1$.

By assumption, the symmetric matrix M has m distinct real eigenvalues and hence orthonormal eigenvectors described by Equation (12). These, in turn, may be used in concert with Equation (11) to yield the desired special unperturbed eigenvectors alluded to above.

Now that $\{y_i^{(0)}\}_{i=1}^m$ are fully determined, we have by Equation (6) the identities

$$\lambda_i^{(1)} = \langle y_i^{(0)}, M^{(1)} y_i^{(0)} \rangle \; (i = 1,\ldots,m). \tag{15}$$

Furthermore, the combination of Equations (12) and (14) yield

$$\langle y_i^{(0)}, M^{(1)} y_j^{(0)} \rangle = 0 \; (i \neq j). \tag{16}$$

The remaining eigenvalue corrections, $\lambda_i^{(k)}$ $(k \geq 2)$, may be obtained from the generalized Dalgarno-Stewart identities.

Whenever Equation (5) is solvable, we will express its solution as

$$y_i^{(k)} = \hat{y}_i^{(k)} + \beta_{1,k}^{(i)} y_1^{(0)} + \beta_{2,k}^{(i)} y_2^{(0)} + \cdots + \beta_{m,k}^{(i)} y_m^{(0)} \; (i = 1,\ldots,m) \tag{17}$$

where $\hat{y}_i^{(k)} := (A_0 - \lambda^{(0)} B_0)^\dagger [-(A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(k-1)} + \sum_{j=0}^{k-2} (\lambda_i^{(k-j)} B_0 + \lambda_i^{(k-j-1)} B_1) y_i^{(j)}]$ has no components in the $\{y_j^{(0)}\}_{j=1}^m$ directions. In light of intermediate normalization, we have $\beta_{i,k}^{(i)} = 0$ $(i = 1,\ldots,m)$. Furthermore, $\beta_{j,k}^{(i)}$ $(i \neq j)$ are to be determined from the condition that Equation (5) be solvable for $k \leftarrow k+1$ and $i = 1,\ldots,m$.

Since, by design, Equation (5) is solvable for $k=1$, we may proceed recursively. After considerable algebraic manipulation, the end result is

$$\beta_{j,k}^{(i)} = \frac{\langle y_j^{(0)}, M^{(1)}\hat{y}_i^{(k)}\rangle - \sum_{l=1}^{k-1}\lambda_i^{(k-l+1)}\beta_{j,l}^{(i)} - \sum_{l=0}^{k-1}\lambda_i^{(k-l)}\langle y_j^{(0)}, B_1 y_i^{(l)}\rangle}{\lambda_i^{(1)} - \lambda_j^{(1)}} \; (i \neq j). \quad (18)$$

The existence of this formula guarantees that each $y_i^{(k)}$ is uniquely determined by enforcing solvability of Equation (5) for $k \leftarrow k+1$.

3.2. Nth Order Degeneracy

We now consider the case of Nth order degeneracy which is characterized by the conditions $\lambda_1^{(j)} = \lambda_2^{(j)} = \cdots = \lambda_m^{(j)} = \lambda^{(j)}$ $(j = 0,\ldots,N-1)$ while $\lambda_i^{(N)}$ $(i = 1,\ldots,m)$ are all distinct. Thus, even though $\lambda^{(j)}$ $(j = 0,\ldots,N-1)$ are determinate, $\{y_i^{(0)}\}_{i=1}^m$ are still indeterminate after enforcing solvability of Equation (5) for $k = N-1$.

Hence, we will determine $\{\lambda_i^{(N)};\, y_i^{(0)}\}_{i=1}^m$ by insisting that Equation (5) be solvable for $k = N$ and $i = 1,\ldots,m$. This requirement is equivalent to the condition that, for each fixed i,

$$\langle x_\mu^{(0)}, -(A_1 - \lambda^{(1)}B_0 - \lambda^{(0)}B_1)y_i^{(N-1)} + (\lambda^{(2)}B_0 + \lambda^{(1)}B_1)y_i^{(N-2)} + \cdots$$
$$+(\lambda_i^{(N)}B_0 + \lambda^{(N-1)}B_1)y_i^{(0)}\rangle = 0 \; (\mu = 1,\ldots,m). \quad (19)$$

Inserting Equation (11) as well as Equation (17) with $k = 1,\ldots,N-1$ and invoking the B_0-orthonormality of $\{x_\mu^{(0)}\}_{\mu=1}^m$, we arrive at, in matrix form,

$$\begin{bmatrix} \langle x_1^{(0)}, M^{(N)}x_1^{(0)}\rangle & \cdots & \langle x_1^{(0)}, M^{(N)}x_m^{(0)}\rangle \\ \vdots & \ddots & \vdots \\ \langle x_m^{(0)}, M^{(N)}x_1^{(0)}\rangle & \cdots & \langle x_m^{(0)}, M^{(N)}x_m^{(0)}\rangle \end{bmatrix} \begin{bmatrix} a_1^{(i)} \\ \vdots \\ a_m^{(i)} \end{bmatrix} = \lambda_i^{(N)} \begin{bmatrix} a_1^{(i)} \\ \vdots \\ a_m^{(i)} \end{bmatrix}, \quad (20)$$

where $M^{(N)}$ is specified by the recurrence relation:

$$M^{(1)} = A_1 - \lambda^{(0)}B_1, \quad (21)$$

$$M^{(N)} = (\lambda^{(N-1)}B_0 - M^{(N-1)})(A_0 - \lambda^{(0)}B_0)^\dagger(A_1 - \lambda^{(1)}B_0 - \lambda^{(0)}B_1)$$
$$- \sum_{l=2}^{N-1}(\lambda^{(N-l)}B_0 - M^{(N-l)})(A_0 - \lambda^{(0)}B_0)^\dagger(\lambda^{(l)}B_0 + \lambda^{(l-1)}B_1) - \lambda^{(N-1)}B_1$$
$$(N = 2,3,\ldots). \quad (22)$$

Thus, each $\lambda_i^{(N)}$ is an eigenvalue with corresponding eigenvector $[a_1^{(i)},\ldots,a_m^{(i)}]^T$ of the matrix M defined by $M_{\mu,\nu} = \langle x_\mu^{(0)}, M^{(N)}x_\nu^{(0)}\rangle$ $(\mu,\nu = 1,\ldots,m)$. It is important to note that, while this recurrence relation guarantees that $\{\lambda_i^{(N)};\ y_i^{(0)}\}_{i=1}^m$ are well defined by enforcing solvability of Equation (5) for $k = N$, $M^{(N)}$ need not be explicitly computed.

By assumption, the symmetric matrix M has m distinct real eigenvalues and hence orthonormal eigenvectors described by Equation (12). These, in turn, may be used in concert with Equation (11) to yield the desired special unperturbed eigenvectors alluded to above.

Now that $\{y_i^{(0)}\}_{i=1}^m$ are fully determined, we have by the combination of Equations (12) and (20) the identities

$$\langle y_i^{(0)}, M^{(N)} y_j^{(0)}\rangle = \lambda_i^{(N)} \cdot \delta_{i,j}. \tag{23}$$

The remaining eigenvalue corrections, $\lambda_i^{(k)}$ $(k \geq N+1)$, may be obtained from the generalized Dalgarno-Stewart identities.

Analogous to the case of first order degeneracy, $\beta_{j,k}^{(i)}$ $(i \neq j)$ of Equation (17) are to be determined from the condition that Equation (5) be solvable for $k \leftarrow k+N$ and $i = 1,\dots,m$. Since, by design, Equation (5) is solvable for $k = 1,\dots,N$, we may proceed recursively. After considerable algebraic manipulation, the end result is

$$\beta_{j,k}^{(i)} = \frac{\langle y_j^{(0)}, M^{(N)} \hat{y}_i^{(k)}\rangle - \sum_{l=1}^{k-1} \lambda_i^{(k-l+N)} \beta_{j,l}^{(i)} - \sum_{l=0}^{k+N-2} \lambda_i^{(k-l+N-1)} \langle y_j^{(0)}, B_1 y_i^{(l)}\rangle}{\lambda_i^{(N)} - \lambda_j^{(N)}} \quad (i \neq j). \tag{24}$$

The existence of this formula guarantees that each $y_i^{(k)}$ is uniquely determined by enforcing solvability of Equation (5) for $k \leftarrow k+N$.

3.3. Mixed Degeneracy

Finally, we arrive at the most general case of mixed degeneracy wherein a degeneracy (multiple eigenvalue) is partially resolved at more than a single order. The analysis expounded upon in the previous sections comprises the core of the procedure for the complete resolution of mixed degeneracy. The following modifications suffice.

During the Rayleigh-Schrödinger procedure, whenever an eigenvalue branches by reduction in multiplicty at any order, one simply replaces the x_μ of Equation (20) by any convenient B_0-orthonormal basis z_μ for the reduced eigenspace. Of course, this new basis is composed of some *a priori* unknown linear combination of the original basis. Equation (24) will still be valid where N is the order of correction where the degeneracy between λ_i and λ_j is resolved. Thus, in general, if λ_i is degenerate to Nth order then $y_i^{(k)}$ will be fully determined by enforcing the solvability of Equation (5) with $k \leftarrow k+N$.

We now present an example which illustrates the general procedure. This example features a simple (i.e. nondegenerate) eigenvalue together with a triple eigenvalue which branches into a single first order degenerate eigenvalue together with a pair of second order degenerate eigenvalues.

4. Example

Define

$$A_0 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, A_1 = \begin{bmatrix} 5 & 3 & 0 & 3 \\ 3 & 5 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 3 & 1 & 0 & 0 \end{bmatrix}, B_0 = \begin{bmatrix} 5 & 3 & 0 & 0 \\ 3 & 5 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}, B_1 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

Using MATLAB's Symbolic Toolbox, we find that

$$\lambda_1(\varepsilon) = \varepsilon \ , \ \lambda_2(\varepsilon) = \varepsilon - \varepsilon^2 - \varepsilon^3 + 2\varepsilon^5 + \cdots \ ,$$

$$\lambda_3(\varepsilon) = 0 \ , \ \lambda_4(\varepsilon) = 1 + \varepsilon^2 + \varepsilon^3 - 2\varepsilon^5 + \cdots \ .$$

Applying the nondegenerate Rayleigh-Schrödinger procedure developed above to

$$\lambda_4^{(0)} = 1; \ x_4^{(0)} = \begin{bmatrix} 0 & 0 & 0 & 1/\sqrt{2} \end{bmatrix}^T,$$

we arrive at

$$\lambda_4^{(1)} = \langle x_4^{(0)}, (A_1 - \lambda_4^{(0)} B_1) x_4^{(0)} \rangle = 0.$$

Solving

$$(A_0 - \lambda_4^{(0)} B_0) x_4^{(1)} = -(A_1 - \lambda_4^{(1)} B_0 - \lambda_4^{(0)} B_1) x_4^{(0)}$$

produces

$$x_4^{(1)} = \begin{bmatrix} 3/4\sqrt{2} & -1/4\sqrt{2} & 0 & 0 \end{bmatrix}^T,$$

where we have enforced the intermediate normalization $\langle x_4^{(1)}, B_0 x_4^{(0)} \rangle = 0$.

In turn, this produces

$$\lambda_4^{(2)} = \langle x_4^{(0)}, (A_1 - \lambda_4^{(0)} B_1) x_4^{(1)} \rangle - \lambda_4^{(1)} \langle x_4^{(0)}, B_1 x_4^{(0)} \rangle = 1,$$

while the generalized Dalgarno-Stewart identities yield

$$\lambda_4^{(3)} = 1.$$

Solving

$$(A_0 - \lambda_4^{(0)} B_0) x_4^{(2)} = -(A_1 - \lambda_4^{(1)} B_0 - \lambda_4^{(0)} B_1) x_4^{(1)} + (\lambda_4^{(2)} B_0 + \lambda_4^{(1)} B_1) x_4^{(0)}$$

produces

$$x_4^{(2)} = \begin{bmatrix} 3/4\sqrt{2} & -1/4\sqrt{2} & 0 & 0 \end{bmatrix}^T,$$

where we have enforced the intermediate normalization $\langle x_4^{(2)}, B_0 x_4^{(0)} \rangle = 0$.

Again, the generalized Dalgarno-Stewart identities yield

$$\lambda_4^{(4)} = 0,\ \lambda_4^{(5)} = -2.$$

We now turn to the mixed degeneracy amongst $\lambda_1^{(0)} = \lambda_2^{(0)} = \lambda_3^{(0)} = \lambda^{(0)} = 0$. With the choice

$$x_1^{(0)} = \begin{bmatrix} 3/4\sqrt{2} & -1/4\sqrt{2} & 0 & 0 \end{bmatrix}^T;\ x_2^{(0)} = \begin{bmatrix} -1/4\sqrt{2} & 3/4\sqrt{2} & 0 & 0 \end{bmatrix}^T;$$

$$x_3^{(0)} = \begin{bmatrix} 0 & 0 & 1/\sqrt{2} & 0 \end{bmatrix}^T,$$

we have from Equation (14), which enforces solvability of Equation (5) for $k = 1$,

$$M = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

with eigenvalues $\lambda_1^{(1)} = \lambda_2^{(1)} = \lambda^{(1)} = 1,\ \lambda_3^{(1)} = 0$.

Thus, $y_1^{(0)}$ and $y_2^{(0)}$ are indeterminate while

$$\begin{bmatrix} a_1^{(3)} & a_2^{(3)} & a_3^{(3)} \end{bmatrix}^T = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T \Rightarrow y_3^{(0)} = \begin{bmatrix} 0 & 0 & 1/\sqrt{2} & 0 \end{bmatrix}^T.$$

Introducing the new basis

$$z_1^{(0)} = \begin{bmatrix} \sqrt{5}/4 & -3/4\sqrt{5} & 0 & 0 \end{bmatrix}^T;\ z_2^{(0)} = \begin{bmatrix} 0 & 1/\sqrt{5} & 0 & 0 \end{bmatrix}^T,$$

we now seek $y_1^{(0)}$ and $y_2^{(0)}$ in the form

$$y_1^{(0)} = b_1^{(1)} z_1^{(0)} + b_2^{(1)} z_2^{(0)};\ y_2^{(0)} = b_1^{(2)} z_1^{(0)} + b_2^{(2)} z_2^{(0)},$$

with orthonormal $\{[b_1^{(1)}, b_2^{(1)}]^T,\ [b_1^{(2)}, b_2^{(2)}]^T\}$.

Solving Equation (5) for $k = 1$,

$$(A_0 - \lambda^{(0)} B_0) y_i^{(1)} = -(A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(0)}\ (i = 1, 2, 3),$$

produces

$$y_1^{(1)} = \begin{bmatrix} \alpha_1 & \beta_1 & \gamma_1 & -(3b_1^{(1)} + b_2^{(1)})/2\sqrt{5} \end{bmatrix}^T;\ y_2^{(1)} = \begin{bmatrix} \alpha_2 & \beta_2 & \gamma_2 & -(3b_1^{(2)} + b_2^{(2)})/2\sqrt{5} \end{bmatrix}^T;$$

$$y_3^{(1)} = \begin{bmatrix} \alpha_3 & \beta_3 & \gamma_3 & 0 \end{bmatrix}^T.$$

Now, enforcing solvability of Equation (5) for $k = 2$,

$$-(A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(1)} + (\lambda_i^{(2)} B_0 + \lambda_i^{(1)} B_1) y_i^{(0)} \perp \{z_1^{(0)}, z_2^{(0)}, y_3^{(0)}\}\ (i = 1, 2, 3),$$

we arrive at

$$M = \begin{bmatrix} -9/10 & -3/10 \\ -3/10 & -1/10 \end{bmatrix},$$

with eigenpairs

$$\lambda_1^{(2)} = 0, \begin{bmatrix} b_1^{(1)} & b_2^{(1)} \end{bmatrix}^T = \begin{bmatrix} 1/\sqrt{10} & -3/\sqrt{10} \end{bmatrix}^T;$$

$$\lambda_2^{(2)} = -1, \begin{bmatrix} b_1^{(2)} & b_2^{(2)} \end{bmatrix}^T = \begin{bmatrix} 3/\sqrt{10} & 1/\sqrt{10} \end{bmatrix}^T \Rightarrow$$

$$y_1^{(0)} = \begin{bmatrix} 1/4\sqrt{2} & -3/4\sqrt{2} & 0 & 0 \end{bmatrix}^T; y_2^{(0)} = \begin{bmatrix} 3/4\sqrt{2} & -1/4\sqrt{2} & 0 & 0 \end{bmatrix}^T,$$

and

$$y_1^{(1)} = \begin{bmatrix} -3\beta_1 & \beta_1 & 0 & 0 \end{bmatrix}^T; y_2^{(1)} = \begin{bmatrix} \alpha_2 & -3\alpha_2 & 0 & -1/\sqrt{2} \end{bmatrix}^T; y_3^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T,$$

as well as $\lambda_3^{(2)} = 0$, where we have invoked intermediate normalization. Observe that $y_1^{(1)}$ and $y_2^{(1)}$ have not yet been fully determined while $y_3^{(1)}$ has indeed been completely specified.

Solving Equation (5) for $k = 2$,

$$(A_0 - \lambda^{(0)} B_0) y_i^{(2)} = -(A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(1)} + (\lambda_i^{(2)} B_0 + \lambda_i^{(1)} B_1) y_i^{(0)} \; (i = 1, 2, 3),$$

produces

$$y_1^{(2)} = \begin{bmatrix} -3b_1 & b_1 & c_1 & 4\beta_1 \end{bmatrix}^T; y_2^{(2)} = \begin{bmatrix} a_2 & -3a_2 & c_2 & -1/\sqrt{2} \end{bmatrix}^T; y_3^{(2)} = \begin{bmatrix} a_3 & b_3 & 0 & 0 \end{bmatrix}^T,$$

where we have invoked intermediate normalization.

We next enforce solvability of Equation (5) for $k = 3$,

$$\langle y_j^{(0)}, -(A_1 - \lambda_i^{(1)} B_0 - \lambda^{(0)} B_1) y_i^{(2)} + (\lambda_i^{(2)} B_0 + \lambda_i^{(1)} B_1) y_i^{(1)} + (\lambda_i^{(3)} B_0 + \lambda_i^{(2)} B_1) y_i^{(0)} \rangle = 0 \; (i \neq j),$$

thereby producing

$$y_1^{(1)} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T; y_2^{(1)} = \begin{bmatrix} 0 & 0 & 0 & -1/\sqrt{2} \end{bmatrix}^T,$$

and

$$y_1^{(2)} = \begin{bmatrix} -3b_1 & b_1 & 0 & 0 \end{bmatrix}^T; y_2^{(2)} = \begin{bmatrix} a_2 & -3a_2 & 0 & -1/\sqrt{2} \end{bmatrix}^T; y_3^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T.$$

With $y_i^{(1)}$ $(i = 1, 2, 3)$ now fully determined, the generalized Dalgarno-Stewart identities yield

$$\lambda_1^{(3)} = 0, \; \lambda_2^{(3)} = -1, \; \lambda_3^{(3)} = 0.$$

Solving Equation (5) for $k = 3$,

$$\begin{aligned}(A_0 - \lambda^{(0)}B_0)y_i^{(3)} &= -(A_1 - \lambda^{(1)}B_0 - \lambda^{(0)}B_1)y_i^{(2)} + (\lambda_i^{(2)}B_0 + \lambda^{(1)}B_1)y_i^{(1)} \\ &\quad + (\lambda_i^{(3)}B_0 + \lambda_i^{(2)}B_1)y_i^{(0)} \ (i = 1,2),\end{aligned}$$

produces

$$y_1^{(3)} = \begin{bmatrix} -3v_1 & v_1 & w_1 & 4b_1 \end{bmatrix}^T; \ y_2^{(3)} = \begin{bmatrix} u_2 & -3u_2 & w_2 & 0 \end{bmatrix}^T,$$

where we have invoked intermediate normalization.

We now enforce solvability of Equation (5) for $k = 4$,

$$\begin{aligned}&\langle y_j^{(0)}, -(A_1 - \lambda^{(1)}B_0 - \lambda^{(0)}B_1)y_i^{(3)} + (\lambda_i^{(2)}B_0 + \lambda^{(1))}B_1)y_i^{(2)} \\ &+ (\lambda_i^{(3)}B_0 + \lambda_i^{(2)}B_1)y_i^{(1)} + (\lambda_i^{(4)}B_0 + \lambda_i^{(3)}B_1)y_i^{(0)}\rangle = 0 \ (i \neq j),\end{aligned}$$

thereby fully determining

$$y_1^{(2)} = \begin{bmatrix} 0 & 0 & 0 & 0 \end{bmatrix}^T; \ y_2^{(2)} = \begin{bmatrix} 0 & 0 & 0 & -1/\sqrt{2} \end{bmatrix}^T.$$

Subsequent application of the generalized Dalgarno-Stewart identities yield

$$\lambda_1^{(4)} = 0, \ \lambda_1^{(5)} = 0, \ \lambda_2^{(4)} = 0, \ \lambda_2^{(5)} = 2, \ \lambda_3^{(4)} = 0, \ \lambda_3^{(5)} = 0.$$

5. Conclusion

In this paper, we have endeavored to provide a comprehensive and unified account of the Rayleigh-Schrödinger perturbation theory for the symmetric definite generalized eigenvalue problem. The cornerstone of our development has been the Moore-Penrose pseudoinverse. Not only does this approach permit a direct analysis of the properties of this procedure but it also obviates the need of alternative approaches for the computation of all of the eigenvectors of the unperturbed matrix pair. Instead, we only require the unperturbed eigenvectors corresponding to those eigenvalues of interest. We have also generalized the Dalgarno-Stewart identities from the standard to the generalized eigenvalue problem.

The focal point of this investigation has been the degenerate case. In light of the inherent complexity of this topic, we have built up the theory gradually with the expectation that the reader would thence not be swept away in a torrent of formulae. We concluded with an example illustrating the general procedure. Observe that the example was worked through *without* explicit computation of the pseudoinverse!

References

[1] Lord Rayleigh, The Theory of Sound, Vol. 1, pp. 113-115, Dover, 1894.

[2] E. Schrödinger, “Quantisierung als Eigenwertproblem”, *Annalen der Physik,* Vol. 80, pp. 437-490, 1926.

[3] B. J. McCartin, “Pseudoinverse Formulation of Rayleigh-Schrödinger Perturbation Theory for the Symmetric Matrix Eigenvalue Problem”, *Journal of Applied Mathematics,* Vol. 2003, No. 9, pp. 459-485, 2003.

[4] Y. Saad, Numerical Methods for Large Eigenvalue Problems, Halsted, 1992.

[5] F. W. Byron, Jr. and R. W. Fuller, Mathematics of Classical and Quantum Physics, pp. 192-205, Dover, 1992.

[6] J. O. Hirschfelder, “Formal Rayleigh-Schrödinger Perturbation Theory for Both Degenerate and Non-Degenerate Energy States”, *International Journal of Quantum Chemistry,* Vol. III, pp. 731-748, 1969.

[7] G. W. Stewart and J. Sun, Matrix Perturbation Theory, Academic, 1990.

[8] J. H. Wilkinson, The Algebraic Eigenvalue Problem, Oxford, 1965.

[9] R. A. Horn and C. R. Johnson, Matrix Analysis, Cambridge, 1990.

[10] B. N. Parlett, The Symmetric Eigenvalue Problem, SIAM, 1998.

[11] M. Konstantinov, V. Mehrmann, and P. Petkov, “Perturbed Spectra of Defective Matrices”, *Journal of Applied Mathematics,* Vol. 2003, No. 3, pp. 115-140, 2003.

[12] F. Rellich, Perturbation Theory of Eigenvalue Problems, Institute of Mathematical Sciences, New York University, 1953.

[13] K. O. Friedrichs, Perturbation of Spectra in Hilbert Space, pp. 90-92, American Mathematical Society, 1965.

[14] T. Kato, Perturbation Theory for Linear Operators, Springer-Verlag, 1980.

[15] A. Dalgarno and A. L. Stewart, “On the perturbation theory of small disturbances”, *Proceedings of the Royal Society of London,* Series A, No. 238, pp. 269-275, 1956.