arXiv:2002.05001v2 [math.SP] 24 Apr 2022

A Tutorial on Matrix Perturbation Theory (using compact matrix notation)

Bassam Bamieh*

Abstract

Analytic perturbation theory for matrices and operators is an immensely useful mathematical technique. Most elementary introductions to this method have their background in the physics literature, and quantum mechanics in particular. In this note, we give an introduction to this method that is independent of any physics notions, and relies purely on concepts from linear algebra. An additional feature of this presentation is that matrix notation and methods are used throughout. In particular, we formulate the equations for each term in the analytic expansions of eigenvalues and eigenvectors as *matrix equations*, namely Sylvester equations in particular. Solvability conditions and explicit expressions for solutions of such matrix equations are given, and expressions for each term in the analytic expansions are given in terms of those solutions. This unified treatment simplifies somewhat the complex notation that is commonly seen in the literature, and in particular, provides relatively compact expressions for the non-Hermitian and degenerate cases, as well as for higher order terms.

1 Introduction

We want to study the behavior of eigenvalues and eigenvectors of matrices that are a function of a "small" parameter ϵ of the form

$$A_\epsilon \;=\; A_0 \;+\; \epsilon A_1, \tag{1}$$

where A_0 and A_1 are given matrices. If the eigenvectors and eigenvalues are analytic functions of ϵ in a neighborhood of zero, then we can write the eigenvalue/eigenvector relations as power series in ϵ, equate terms of same powers in ϵ and derive expressions for each set of terms. These expressions and the related algebra can get messy. It is shown in this document that by adopting matrix notation, expressions are simplified and compactified, and additional insight is obtained. It is rather ironic that *matrix notation* is not fully utilized in standard treatments of *matrix* perturbation theory. It can be argued that a better way to think about finding expansion terms in the eigenvectors is to treat them all together as a matrix, rather than as individual vectors. Better insight is achieved in this manner, especially for the cases of degenerate eigenvalues and higher order terms.

Notation and Preliminaries

Before we begin, we set some useful matrix notation for eigenvector/eigenvalue relations, as well as manipulations with diagonal matrices and the Hadamard product.

- A vector v_i (w_i^*) is a right (left) eigenvector of a matrix A if

$$Av_i \;=\; \lambda_i v_i, \qquad (w_i^* A \;=\; \lambda w_i^*)\,.$$

*Department of Mechanical Engineering, University of California at Santa Barbara, *bamieh@ucsb.edu*. This work is partially supported by NSF Awards CMMI-1763064 and ECCS-1932777.

Throughout this note, we will assume the semi-simple case, i.e. that A_ϵ has a full set of eigenvectors (i.e. diagonalizable) for each ϵ in some neighborhood of zero. In this case, for an $n \times n$ matrix A, there are n eigenvalue/vector relations

$$Av_i \;=\; \lambda_i v_i \qquad\qquad (w_i^* A \;=\; \lambda_i w_i^*)\,, \qquad\qquad i = 1, \ldots, n,$$

It is very useful to note that these n relations can be compactly rewritten as a matrix equation

$$\begin{bmatrix} Av_1 & \cdots & Av_n \end{bmatrix} = \begin{bmatrix} \lambda_1 v_1 & \cdots & \lambda_n v_n \end{bmatrix}$$

$$\Leftrightarrow \quad \begin{bmatrix} \\ A \\ \\ \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \quad \Leftrightarrow \quad AV = V\Lambda, \tag{2}$$

where V is a matrix whose columns are the eigenvectors of A, and Λ is the diagonal matrix made up of the eigenvalues of A. Similarly, we also have

$$W^* A \;=\; \Lambda W^*,$$

where the rows of W^* are the left eigenvectors. Note that in the special case of Hermitian (or more generally normal) matrices, the left and right eigenvectors coincide, i.e. $W = V$.

- For any square matrix M, let $\mathsf{dg}(M)$ be itself a square diagonal matrix made up of only the diagonal elements of M. Clearly $\mathsf{dg}(M+N) \;=\; \mathsf{dg}(M) \;+ \mathsf{dg}(N)$ for any two matrices M and N. If D is a *diagonal* matrix with compatible dimensions, then it immediately follows that $\mathsf{dg}(MD) = \mathsf{dg}(DM) = \mathsf{dg}(M)\,\mathsf{dg}(D)$. Thus for any two matrices M and N with compatible dimensions

$$\mathsf{dg}\Big(M\;\mathsf{dg}(N)\Big) \;\;=\;\; \mathsf{dg}(M)\,\mathsf{dg}(N)\,.$$

- The Hadamard product $M \circ N$ of two matrices is the element-by-element product. It is distributive over matrix additions, but not over matrix products generally except with diagonal matrices

$$M \circ \big(N_1 + N_2\big) \;=\; M \circ N_1 + M \circ N_2, \qquad\qquad M \circ \big(N\Lambda\big) \;=\; \big(M \circ N\big)\Lambda,$$

when Λ is diagonal. If $M = uw^*$ is rank 1, then it follows that

$$M \circ N \;=\; (uw^*) \circ N \;=\; \mathsf{diag}(u)\,N\;\mathsf{diag}(w)\,,$$

where $\mathsf{diag}(u)$ is the diagonal matrix formed from the entries of the vector u. Thus if $M = \sum_i \lambda_i u_i w_i^*$ is a diagonalizable matrix, then we can obtain an expression for the Hadamard product in terms of a sum of standard matrix products

$$M \circ N \;=\; \left(\sum_i \lambda_i u_i w_i^*\right) \circ N \;=\; \sum_i \lambda_i\;\mathsf{diag}(u_i)\,N\;\mathsf{diag}(w_i)$$

We will use the notation $M^{k\circ}$ for the Hadamard k power of M, i.e. the matrix whose entries are m_{ij}^k, with m_{ij} being the entries of M. Similarly, $M^{-\circ}$ is the matrix with entries $1/m_{ij}$ if all entries are $m_{ij} \neq 0$. The Hadamard pseudo-inverse $M^{\dagger\circ}$ is the matrix with entries $1/m_{ij}$ if $m_{ij} \neq 0$, and 0 if $m_{ij} = 0$.

2 Perturbation Expansion

Consider (1) and assume that the eigenvectors and eigenvalues are analytic functions of ϵ in a neighborhood of 0. The eigenvector/eigenvalue relationship then reads as

$$\begin{aligned} A_\epsilon\, V_\epsilon \;&=\; V_\epsilon\, \Lambda_\epsilon, \qquad W_\epsilon^*\, A_\epsilon \;=\; \Lambda_\epsilon W_\epsilon^* \\ &\Updownarrow \\ (A_0 \;+\; \epsilon A_1)\,(V_0 \;+\; \epsilon V_1 \;+\; \cdots) &= \;(V_0 \;+\; \epsilon V_1 \;+\; \cdots)\,(\Lambda_0 \;+\; \epsilon \Lambda_1 \;+\; \cdots) \qquad (3) \\ (W_0^* \;+\; \epsilon W_1^* \;+\; \cdots)\,(A_0 \;+\; \epsilon A_1) &= \;(\Lambda_0 \;+\; \epsilon \Lambda_1 \;+\; \cdots)\,(W_0^* \;+\; \epsilon W_1^* \;+\; \cdots) \qquad (4) \end{aligned}$$

Note that Λ and Λ_i's are *diagonal matrices*. These equations describe eigenvectors that each belong to a one dimensional subspace, and thus V_ϵ and W_ϵ are not unique unless we impose some normalization constraint. There are many possible such constraints. The first we will use is the reciprocal basis constraint $W^*(\epsilon)V(\epsilon) = I$ which gives

$$W^*(\epsilon)V(\epsilon) = I \qquad \Leftrightarrow \qquad (W_0^* \;+\; \epsilon W_1^* \;+\; \cdots)\,(V_0 \;+\; \epsilon V_1 \;+\; \cdots) \;=\; I. \qquad (5)$$

Equating equal powers of ϵ in (3) gives a sequence of matrix equations

$$\begin{aligned} A_0V_0 &= V_0\Lambda_0 & W_0^*A_0 &= \Lambda_0W^* & (6)\\ A_0V_1 + A_1V_0 &= V_0\Lambda_1 + V_1\Lambda_0 & W_1^*A_0 + W_0^*A_1 &= \Lambda_1W_0^* + \Lambda_0W_1^* & (7)\\ A_0V_2 + A_1V_1 &= V_0\Lambda_2 + V_1\Lambda_1 + V_2\Lambda_0 & W_2^*A_0 + W_1^*A_1 &= \Lambda_2W_0^* + \Lambda_1W_1^* + \Lambda_0W_2^* & (8)\\ &\vdots \\ A_0V_k + A_1V_{k-1} &= \sum_{i=0}^{k} V_i\Lambda_{k-i} & W_k^*A_0 + W_{k-1}^*A_1 &= \sum_{i=0}^{k} \Lambda_{k-i}W_i^* & (9) \end{aligned}$$

A similar exercise for (5) gives

$$W_0^*V_0 = I, \qquad W_0^*V_1 + W_1^*V_0 = 0, \qquad \cdots \qquad \sum_{i=0}^{k} W_i^*V_{k-i} = 0. \qquad (10)$$

In the special case of Hermitian matrices, we have $W_\epsilon = V_\epsilon$, and the eigenvectors are thus normalized to be orthonormal. We also then have the condition $V_0^*V_1 = -V_1^*V_0$, i.e. $V_0^*V_1$ is skew-Hermitian, which implies that its diagonal entries must be imaginary, or zero if the vectors are real. In the real case, the diagonal entries are $v_{0i}^*v_{1i} = 0$, i.e. for each i, v_{0i} and v_{1i} are orthogonal. Geometrically, this means that the curve $v_i(\epsilon)$ has a tangent at $\epsilon = 0$ that is orthogonal to the direction of the unperturbed eigenvector v_{0i}. More generally, the normalization condition insures that the curves $v_i(\epsilon)$ lie on the sphere in $\mathbb{R}^n$ for any ϵ (thus the orthogonality of their initial tangents to each v_{0i}).

3 Calculating First order Terms

We first begin by calculating the first order behavior of the eigenvalues. This is the term Λ_1, for which we can use equation (7). Left multiplication by W_0^* (to get rid of the V_0 factor multiplying Λ_1) gives

$$\begin{aligned} W_0^*A_0V_1 + W_0^*A_1V_0 &= W_0^*V_0\Lambda_1 + W_0^*V_1\Lambda_0 \\ \Leftrightarrow \qquad \Lambda_0W_0^*V_1 - W_0^*V_1\Lambda_0 + W_0^*A_1V_0 &= \Lambda_1. \end{aligned}$$

Now make the following observations

- Since Λ_0 is diagonal, then $(W_0^*V_1)\,\Lambda_0$ and $\Lambda_0\,(W_0^*V_1)$ have equal diagonals, i.e. $W_0^*V_1\Lambda_0 - \Lambda_0W_0^*V_1$ has zeros on the diagonal.

- Since Λ_1 must be diagonal, then

$$\Lambda_1 = \mathsf{dg}\Big(W_0^* A_1 V_0 - W_0^* V_1 \Lambda_0 + \Lambda_0 W_0^* V_1\Big) = \mathsf{dg}\Big(W_0^* A_1 V_0\Big) + 0$$
$$\Rightarrow \quad \boxed{\Lambda_1 = \mathsf{dg}\Big(W_0^* A_1 V_0\Big)} \tag{11}$$

Thus the first order correction to the i'th eigenvalue is $\lambda_{1i} = w_{0i}^* A_1 v_{0i}$, which is the well known expression[1]. To fully appreciate this, expand (11) using partitioned matrix notation

$$\Lambda_1 = \mathsf{dg}\left(\begin{bmatrix} w_{01}^* \\ \vdots \\ w_{0n}^* \end{bmatrix} \begin{bmatrix} A \end{bmatrix} \begin{bmatrix} v_{01} & \cdots & v_{0n} \end{bmatrix}\right) = \begin{bmatrix} (w_{01}^* A_1 v_{01}) & & \\ & \ddots & \\ & & (w_{0n}^* A_1 v_{0n}) \end{bmatrix}.$$

3.1 Calculating First Order Eigenvector Terms: Distinct Eigenvalues

Equations (7) can be rewritten as matrix equations with V_1 and W_1 as the unknowns respectively as follows

$$A_0 V_1 + A_1 V_0 = V_0 \Lambda_1 + V_1 \Lambda_0 \quad \Leftrightarrow \quad A_0 V_1 - V_1 \Lambda_0 = (V_0 \Lambda_1 - A_1 V_0), \tag{12}$$
$$W_1^* A_0 - \Lambda_0 W_1^* = (\Lambda_1 W_0^* - W_0^* A_1) \tag{13}$$

The last two equations are Sylvester equations for the matrices V_1 and W_1 respectively. We first discuss (12). Define the matrix-valued operator $\mathcal{L}(X) := A_0 X - X \Lambda_0$, and note that (12) can be rewritten as $\mathcal{L}(V_1) = (V_0 \Lambda_1 - A_1 V_0)$. The properties of this "Sylvester operator" are discussed in Appendix A.1, and the solvability of equation (12) is determined by those properties. The following matrices appear in the solution: V_0, W_0, the eigenvectors of A_0, A_0^* respectively, and the matrices whose ij'th entries are given by

$$\big(\mathbf{\Pi}\big)_{ij} := \lambda_{0i} - \lambda_{0j}, \qquad \big(\mathbf{\Pi}^{\dagger\circ}\big)_{ij} := \begin{cases} \dfrac{1}{\lambda_{0i} - \lambda_{0j}}, & i \neq j, \\ 0, & i = j. \end{cases}$$

Entries of $\mathbf{\Pi}$ are made up of all possible sums of the eigenvalues of A_0 and $-\Lambda_0$ (i.e. differences of the eigenvalues of A_0), and $\mathbf{\Pi}^{\dagger\circ}$ is the Hadamard (element-by-element) pseudo-inverse of $\mathbf{\Pi}$. We now apply the machinery in the Appendices to the particular operator $\mathcal{L}$ in equation (12).

- The operator $\mathcal{L}$ has the following spectral decomposition (applying formula (24) with $U = Z = I$)

$$\mathcal{L}(X) = V_0 \Big(\mathbf{\Pi} \circ \big(W_0^* X\big)\Big) = \sum_{ij} (\lambda_{0i} - \lambda_{0j}) \, \big(w_{0i}^* X e_{0j}\big) \, v_{0i} e_j^*. \tag{14}$$

  Since for $i = j$, $\lambda_{0i} - \lambda_{0j} = 0$, $\mathcal{L}$ is rank deficient by at least n.

- Even though $\mathcal{L}$ is not of full rank, equation (12) is solvable since its right hand side is in the range of $\mathcal{L}$. This argument is detailed in Appendix A.2.

- The minimum norm solution of (12) is obtained from the formula (eq. (26), where we again use $U = Z = I$) for the pseudo-inverse $\mathcal{L}^\dagger$

$$V_1 = \mathcal{L}^\dagger (V_0 \Lambda_1 - A_1 V_0) = V_0 \Big(\mathbf{\Pi}^{\dagger\circ} \circ \big(W_0^* (V_0 \Lambda_1 - A_1 V_0)\big)\Big) = V_0 \Big(\mathbf{\Pi}^{\dagger\circ} \circ \big(\Lambda_1 - W_0^* A_1 V_0\big)\Big)$$
$$\Rightarrow \quad \boxed{V_1 = -V_0 \Big(\mathbf{\Pi}^{\dagger\circ} \circ \big(W_0^* A_1 V_0\big)\Big)}, \tag{15}$$

  where we used $W_0^* V_0 = I$, the distributive property of the Hadamard product $\circ$, and the fact that $\mathbf{\Pi}^{\dagger\circ} \Lambda_1 = 0$ since Λ_1 is diagonal and $\mathbf{\Pi}^{\dagger\circ}$ has zeros on the diagonal.

- The Sylvester operator in (13) is just $\mathcal{L}^*$, and we can similarly derive (see Appendix A.2) the minimum-norm solution

$$\boxed{W_1^* \;=\; \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) W_0^*}\,.$$

To compare the solution formula (15) with standard expressions in the literature, we expand it as

$$V_1 \;=\; -\,V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) \;=\; -\sum_{i\neq j} \frac{1}{\lambda_{0i}-\lambda_{0j}} \left(w_{0i}^* A_1 v_{0j}\right)\, v_{0i} e_j^*$$

Note that V_1 is a matrix whose k'th column is v_{1k}. The k'th column is obtained from $V_1 e_k$

$$v_{1k} \;=\; -\,V_1 e_k \;=\; -\sum_{i\neq j} \frac{w_{0i}^* A_1 v_{0j}}{\lambda_{0i}-\lambda_{0j}}\, v_{0i} e_j^*\, e_k \;=\; -\sum_{i\neq k} \frac{w_{0i}^* A_1 v_{0k}}{\lambda_{0i}-\lambda_{0k}}\, v_{0i}, \;=\; \sum_{i\neq k} \frac{w_{0i}^* A_1 v_{0k}}{\lambda_{0k}-\lambda_{0i}}\, v_{0i}, \tag{16}$$

which is the standard expression in the literature [1].

Uniqueness of V_1 and W_1

Note that (15) is just one solution to the Sylvester equation. All other solutions can be obtained by adding arbitrary elements of the null space of $\mathcal{L}$. In Appendix A.2, it is shown how the null spaces of $\mathcal{L}$ and $\mathcal{L}^*$ are characterized, and from that we can conclude that all solutions of the Sylvester equation (12) and (13) can be written as

$$V_1 \;=\; -\,V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) + V_0 D_1, \qquad W_1^* \;=\; \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) W_0^* - D_2 W_0^*, \tag{17}$$

where D_1 and D_2 are arbitrary diagonal matrices. The normalization conditions (10) now give

$$\begin{aligned} 0 \;=\; W_0^* V_1 + W_1^* V_0 = \;& -\,W_0^* V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) + \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) W_0^* V_0 + D_1 - D_2 \\ =\; & D_1 - D_2. \end{aligned} \tag{18}$$

We now make several observations

- In the self-adjoint case, we have $W_0 = V_0$ and $W_1 = V_1$, and the normalization conditions (10) give

  $$0 \;=\; V_0^* V_1 \;=\; V_0^* \left(-V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) + V_0 D_1\right) \;=\; -\,\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right) + D_1.$$

  Since the first term has zeros on the diagonal ($\mathbf{\Pi}^{\dagger\circ}$ is zero on the diagonal) and D_1 is diagonal, we conclude that $D_1 = 0$. Thus, the minimum norm solution (15) to the Sylvester equation does indeed satisfy the normalization conditions (10).

- In the general (non self-adjoint) case, it seems that the following set of solutions

  $$V_1 \;=\; -\,V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) + V_0 D, \qquad W_1^* \;=\; \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) W_0^* - D W_0^*$$

  where D is any diagonal matrix will all satisfy the normalization condition (10) up to first order. Since V_1 and W_1 must be unique[1], it seems like consideration of higher order terms is necessary to find the correctly normalized solutions.

[1] The condition $W_\epsilon^* V_\epsilon = I$ uniquely determine the functions V_ϵ and W_ϵ, and therefore uniquely determines their Taylor series expansion terms $\{V_k\}$ and $\{W_k\}$.

- In some references [2, Eq. 5.1.31], a non-standard normalization is used. In our notation, this normalization is stated as follows

$$\begin{aligned} w_{0i}^* v_{0j} &= \delta_{i-j}, \qquad w_{0i}^* v_{\epsilon i} = 1, \\ W_0^* V_0 &= I, \qquad \mathsf{dg}(W_0^* V_\epsilon) = I \quad \Rightarrow \quad \forall k \geq 1,\ \mathsf{dg}(W_0^* V_k) = 0. \end{aligned} \tag{19}$$

  This normalization will uniquely determine $v_{\epsilon i}$ unless it becomes orthogonal to w_{0i}. For finite matrices, we can guarantee that $v_{\epsilon i}$ will not be orthogonal to w_{0i} for ϵ in some neighborhood of 0.

  This non-standard normalization is used because it simplifies the recursive formulae for higher order peturbation terms as we will see later on. For now, observe that if we adopt this normalization and check the general solutions (17)

$$\begin{aligned} 0 = \mathsf{dg}(W_0^* V_1) &= \mathsf{dg}\Big(-W_0^* V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\right) + W_0^* V_0 D_1\Big) \\ &= -\mathsf{dg}\big(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0\right)\big) + D_1 = 0 + D_1 \end{aligned}$$

  So we can conclude that the minimum-norm solution (15) is indeed normalized with the non-standard normalization (19).

3.2 Calculating First Order Eigenvector Terms V_1: Repeated (Degenerate) Case

We consider the case when the eigenvalues are repeated, but A_0 has a full set of eigenvectors. This condition is automatically satisfied in the Hermitian or normal A_0 case, but may not be always true in the non-normal A_0 case. We will consider the case of normal A_0, A_1 here, for which we have $W_0 = V_0$, and therefore $A_0 = V_0 \Lambda_0 V_0^*$.

The main difficulty when A_0 has repeated eigenvalues is that there is not a unique choice of the eigenvectors V_0 in this case, even after normalization. For any repeated eigenvalue, there corresponds an m-dimensional invariant subspace, where m is the geometric multiplicity of the eigenvalue. Any basis of this invariant subspace is composed of eigenvectors. It turns out that there is a special choice of basis which renders the expansion (3) valid. The main task is to find such a basis.

Now assume A_0 has a repeated eigenvalue $\bar{\lambda}$ (multiplicity m) and define a basis so that A_0 is block-diagonal with the first m basis elements a basis for the eigensubspace of $\bar{\lambda}$. With this basis, A_0, V_0, W_0 and Λ_0 have a block partitioning as

$$A_0 = \begin{bmatrix} (A_0)_{11} & 0 \\ 0 & (A_0)_{22} \end{bmatrix} = \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix} \begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix} \begin{bmatrix} (V_0^*)_{11} & 0 \\ 0 & (V_0^*)_{22} \end{bmatrix} = V_0 \Lambda_0 V_0^*,$$

where I is the $m \times m$ identity matrix. Clearly there is not a unique choice for $(V_0)_{11}$ since replacing it with $(V_0)_{11} U$, where U is any unitary matrix will keep the above expression valid[2].

Now take eq. (7)

$$\Lambda_0 V_1 + A_1 V_0 = V_0 \Lambda_1 + V_1 \Lambda_0.$$

If we partition all matrices conformably with the above partitions, we get

$$\begin{aligned} &\begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix} \begin{bmatrix} (V_1)_{11} & (V_1)_{12} \\ (V_1)_{21} & (V_1)_{22} \end{bmatrix} + \begin{bmatrix} (A_1)_{11} & (A_1)_{12} \\ (A_1)_{21} & (A_1)_{22} \end{bmatrix} \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix} \\ &\qquad = \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix} \begin{bmatrix} (\Lambda_1)_{11} & 0 \\ 0 & (\Lambda_1)_{22} \end{bmatrix} + \begin{bmatrix} (V_1)_{11} & (V_1)_{12} \\ (V_1)_{21} & (V_1)_{22} \end{bmatrix} \begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix}. \end{aligned}$$

Taking the 11 block equation we get

$$\bar{\lambda} I_n\ (V_1)_{11} + (A_1)_{11} (V_0)_{11} = (V_0)_{11} (\Lambda_1)_{11} + (V_1)_{11}\ \bar{\lambda} I_n,$$

[2] U unitary means $UU^* = I$, and therefore $(V_0)_{11} U\ \bar{\lambda} I\ U^* (V_0^*)_{11} = (V_0)_{11}\ \bar{\lambda} I\ (V_0^*)_{11}$.

but $\bar{\lambda} I_n$ commutes with any matrix, so we simply get

$$(A_1)_{11}(V_0)_{11} \;=\; (V_0)_{11}(\Lambda_1)_{11},$$

from which we can obtain $(\Lambda_1)_{11}$ by left multiplication by $(V_0^*)_{11}$

$$(\Lambda_1)_{11} \;=\; (V_0^*)_{11}(A_1)_{11}(V_0)_{11} \;=\; (V_0^* A_1 V_0)_{11},$$

where the last equality follows from the block-diagonal structure of V_0 and V_0^*. Now let's examine the meaning of this. The perturbation expansion (3) assumes that all the Λ_i's are diagonal matrices (otherwise (3) is not an eigenvalue/eigenvector relation). This equation states that for $(\Lambda_1)_{11}$ to be diagonal as required, *we must choose* $(V_0)_{11}$ *so that its columns are the eigenvectors of* $(A_1)_{11}$. Consequently, $(\Lambda_1)_{11}$ will simply be the diagonal matrix of eigenvalues of $(A_1)_{11}$.

The meaning of the above is that only this particular choice of V_0 will yield the expansion (3). Unlike the non-repeated case, we also have to discover "the special" V_0, not just V_1 and Λ_1.

Now to calculate V_1 we can apply the psuedo-inverse formula (26) (which produced (15) earlier), but now we pay special attention to the structure of $\mathbf{\Pi}^{\dagger\circ}$. In this case, it is

$$\left(\mathbf{\Pi}^{\dagger\circ}\right)_{ij} \;:=\; \begin{cases} \dfrac{1}{\lambda_{0i} - \lambda_{0j}}, & \lambda_{0i} \neq \lambda_{0j}, \\ 0, & \lambda_{0i} = \lambda_{0j}. \end{cases}$$

It is instructive to look at the structure of $\mathbf{\Pi}^{\dagger\circ}$, it is of the following form

$$\mathbf{\Pi}^{\dagger\circ} \;=\; \begin{bmatrix} 0 & & \\ & 0 & \\ & & \ddots \end{bmatrix},$$

where the large 0 block is $m \times m$, the diagonal is all zeros, and the remaining terms are $1/(\lambda_{0i} - \lambda_{0j})$. The expression for V_1 is now the same as (15), but now we also make sure to use the V_0 that made Λ_1 diagonal above (recall that the derivation of (15) assumed Λ_1 to be diagonal)

$$V_1 \;=\; -\,V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(V_0^* A_1 V_0\right)\right).$$

Again, to compare with existing expressions, find v_{1k} using $V_1 e_k$

$$v_{1k} \;=\; -\,V_1 e_k \;=\; -\sum_{\lambda_{0i} \neq \lambda_{0k}} \frac{v_{0i}^* A_1 v_{0j}}{\lambda_{0i} - \lambda_{0j}}\; v_{0i} e_j^*\; e_k \;=\; \sum_{\lambda_{0i} \neq \lambda_{0k}} \frac{v_{0i}^* A_1 v_{0k}}{\lambda_{0k} - \lambda_{0i}}\; v_{0i}, \tag{20}$$

Example 1. Consider the simplest case of 2×2 matrices

$$A_0 \;+\; \epsilon A_1 \;:=\; I \;+\; \epsilon M,$$

where M is a normal matrix. $A_0 = I$ has two eigenvalues at 1 with multiplicity 2, and a 2-dimensional eigenspace (all of R^2). Let M have eigenvalue/vector pairs (α, x) and (β, y). We actually know the eigenvalues of $I + \epsilon M$ explicitly as a function of ϵ as (since $I + \epsilon M$ has the same eigenvectors as M)

$$\lambda_{\epsilon 1} \;=\; 1 \;+\; \epsilon\, \alpha, \qquad\qquad \lambda_{\epsilon 2} \;=\; 1 \;+\; \epsilon\, \beta.$$

Let's see if this can be replicated using the procedure described above. We need to choose V_0 so that its columns are eigenvectors of M, so the choice is

$$V_0 \;=\; \left[\; x \;\vdots\; y \;\right],$$

and therefore

$$V_0^* A_1 V_0 \;=\; \left[\; x \;\vdots\; y \;\right]^* M \left[\; x \;\vdots\; y \;\right] \;=\; \left[\begin{array}{c:c} \alpha & 0 \\ \hdashline 0 & \beta \end{array}\right].$$

Note that the off-diagonal elements are zero since the vectors x and y are orthogonal (M is normal).

4 Higher Order Terms

The following normalization will lead to rather compact recursive formulae for terms of all orders

$$\begin{aligned} w_{0i}^* v_{0j} &= \delta_{i-j}, \qquad & w_{0i}^* v_{\epsilon i} &= 1, \\ W_0^* V_0 &= I, \qquad & \mathsf{dg}(W_0^* V_\epsilon) &= I \quad \Rightarrow \quad \forall k \geq 1,\ \mathsf{dg}(W_0^* V_k) = 0. \end{aligned} \tag{21}$$

We find the eigenvalues first. Decompose $\Lambda_\epsilon = (\Lambda_0 + \epsilon\Lambda_1 + \cdots) =: (\Lambda_0 + \tilde{\Lambda})$, and rearrange to obtain an equation for $\tilde{\Lambda}$

$$(A_0 + \epsilon A_1) V_\epsilon = V_\epsilon \left(\Lambda_0 + \tilde{\Lambda}\right) \tag{22}$$

$$\begin{aligned} V_\epsilon \tilde{\Lambda} &= -V_\epsilon \Lambda_0 + A_0 V_\epsilon + \epsilon A_1 V_\epsilon && \text{rearrange terms} \\ W_0^* V_\epsilon \tilde{\Lambda} &= -W_0^* V_\epsilon \Lambda_0 + W_0^* A_0 V_\epsilon + \epsilon W_0^* A_1 V_\epsilon && \text{left multiply by } W_0^* \\ \mathsf{dg}\left(W_0^* V_\epsilon \tilde{\Lambda}\right) &= -\mathsf{dg}(W_0^* V_\epsilon \Lambda_0) + \mathsf{dg}(\Lambda_0\ W_0^* V_\epsilon) + \mathsf{dg}(W_0^* A_1\ \epsilon V_\epsilon) && \text{take diagonals} \\ \mathsf{dg}(W_0^* V_\epsilon)\,\tilde{\Lambda} &= -\mathsf{dg}(W_0^* V_\epsilon)\,\Lambda_0 + \Lambda_0\ \mathsf{dg}(W_0^* V_\epsilon) + \mathsf{dg}(W_0^* A_1\ \epsilon V_\epsilon) && \mathsf{dg}(M\Lambda) = \mathsf{dg}(M)\,\Lambda \text{ if } \Lambda \text{ diagonal} \\ \tilde{\Lambda} &= \mathsf{dg}(W_0^* A_1\ \epsilon V_\epsilon) && \mathsf{dg}(W_0^* V_\epsilon) = I \\ \left(\epsilon\Lambda_1 + \epsilon^2\Lambda_2 + \cdots\right) &= \mathsf{dg}\Big(W_0^* A_1\ \epsilon\big(V_0 + \epsilon V_1 + \cdots\big)\Big) && \text{expand \& equate} \end{aligned}$$

$$\boxed{\Lambda_k = \mathsf{dg}\left(W_0^* A_1 V_{k-1}\right)} \tag{23}$$

The first of these formulae is the familiar $\Lambda_1 = \mathsf{dg}(W_0^* A_1 V_0)$. Successive Λ_k's require finding the vectors V_{k-1} that are normalized according to (21) (and not the standard normalization).

To find V_k's, recall and rearrange equation (9) into a Sylvester equation for V_k

$$\begin{aligned} A_0 V_k + A_1 V_{k-1} &= V_0\Lambda_k + V_1\Lambda_{k-1} \cdots + V_{k-1}\Lambda_1 + V_k\Lambda_0 \\ A_0 V_k - \Lambda_k V_0 &= V_0\Lambda_k + V_1\Lambda_{k-1} + \cdots + V_{k-1}\Lambda_1 - A_1 V_{k-1} \end{aligned}$$

All solutions to this Sylvester equation are given by the psuedo-inverse formula

$$V_k = V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ W_0^* \left(V_0\Lambda_k + V_1\Lambda_{k-1} + \cdots + V_{k-1}\Lambda_1 - A_1 V_{k-1}\right)\right) + V_0 D_k,$$

where D_k is any diagonal matrix. Enforcing the normalization $W_0^* V_0 = I$, $\mathsf{dg}(W_0^* V_k) = 0$, $k \geq 1$ gives that $D_0 = I$ and $D_k = 0$ for $k \geq 1$, and we conclude

$$\boxed{V_k = V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ W_0^* \left(V_0\Lambda_k + V_1\Lambda_{k-1} + \cdots + V_{k-1}\Lambda_1 - A_1 V_{k-1}\right)\right)}.$$

When this formula is combined with (23) for Λ_k, we see that we can obtain V_k using previous calculations for the terms $V_0, \ldots, V_{k-1}$ and $\Lambda_0, \ldots, \Lambda_{k-1}$. For computations, the following equivalent form may be easier to implement

$$V_k = V_0 \left(\sum_{i=0}^{k-1} \left(\mathbf{\Pi}^{\dagger\circ} \circ W_0^* V_i\right)\Lambda_{i-1} - \mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_{k-1}\right)\right).$$

5 Acknowledgement

The author would like to acknowledge helpful input from Maurice Filo, Stacy Patterson and Karthik Chickmaglur.

A Appendix

A.1 Spectral Decomposition of the Sylvester Operator

Consider the Sylvester matrix equation (in X) of the form

$$AX \;+\; XB \;=\; Q,$$

where A, B and Q are given (square) matrices. Solutions and properties of this equation are determined by the Sylvester operator $\mathcal{L}$

$$\mathcal{L}(X) \;:=\; AX \;+\; XB.$$

We will need to find the "eigen-matrices" of $\mathcal{L}$ and its adjoint $\mathcal{L}^*$. First we calculate $\mathcal{L}^*$ using the standard inner product on matrices $\langle M, N\rangle = \mathsf{tr}\big(M^*N\big)$. Starting from the $\equiv$ equality

$$\mathsf{tr}\big(X^*(A^*Y + YB^*)\big) = \mathsf{tr}\big(X^*A^*Y + B^*X^*Y\big) = \mathsf{tr}\big((AX+XB)^*Y\big) = \langle \mathcal{L}(X), Y\rangle \equiv \langle X, \mathcal{L}^*(Y)\rangle = \mathsf{tr}\big(X^*\mathcal{L}^*(Y)\big)$$

we see that

$$\mathcal{L}^*(Y) \;=\; A^*Y + YB^*.$$

Thus in particular, $\mathcal{L}$ is self-adjoint only if A and B are self-adjoint[3]

Now we assume that A and B are diagonalizable (i.e. have a full set of linearly independent eigenvectors each). A spectral decomposition of $\mathcal{L}$ can be obtained from the spectral decompositions of A and B. This will then allow for applying arbitrary analytic functions on $\mathcal{L}$ including inversion.

Denote the eigenvalues and right/left eigenvectors of A and B as follows

$$\begin{array}{rclcrcl} Av_i &=& \lambda_i v_i, & & A^*w_i &=& \lambda_i^* w_i, \\ Bu_j &=& \gamma_j u_j, & & B^*z_j &=& \gamma_j^* z_j, \end{array} \quad \text{or in matrix form} \quad \begin{array}{rclcrcl} AV &=& V\Lambda, & & A^*W &=& W\Lambda^* \\ BU &=& U\Gamma, & & B^*Z &=& Z\Gamma^* \end{array}.$$

Note that $W = V^{-*}$ and $Z = U^{-*}$. The n^2 eigenvalues and "eigen-matrices" of $\mathcal{L}$ and $\mathcal{L}^*$ are found as follows

$$\begin{aligned} \mathcal{L}(v_iz_j^*) &= A\; v_iz_j^* \;+\; v_iz_j^*\; B \;\;\;\; = \lambda_i\; v_iz_j^* \;+\; \gamma_j\; v_iz_j^* \;=\; (\lambda_i+\gamma_j)\; v_iz_j^*, \\ \mathcal{L}^*(w_iu_j^*) &= A^*\; w_iu_j^* \;+\; w_iu_j^*\; B^* = \lambda_i^*\; w_iu_j^* \;+\; \gamma_j^* w_iu_j^* = (\lambda_i^*+\gamma_j^*)\; w_iu_j^*. \end{aligned}$$

In other words, eigenvalues of $\mathcal{L}$ (resp. $\mathcal{L}^*$) are all n^2 possible combinations $\lambda_i + \gamma_j$ (resp. $(\lambda_i+\gamma_j)^*$) of eigenvalues of A and B, and the eigen-matrices are all the corresponding outer products of right/left eigenvectors of A and B.

Using the above, a spectral decomposition of $\mathcal{L}$ can now be written as follows

$$\mathcal{L}(X) \;=\; \sum_{ij} (\lambda_i+\gamma_j)\,\big\langle w_iu_j^*, X\big\rangle\; v_iz_j^*.$$

This can be rewritten in compact notation. First define the matrix $\mathbf{\Pi}$ whose ij'th entry is

$$\big(\mathbf{\Pi}\big)_{ij} \;:=\; \lambda_i + \gamma_j.$$

Noting that $v_i = Ve_i$ (where e_i is the vector of all zeros except 1 in the i'th row), and similarly for the other eigenvectors, calculate

$$\begin{aligned} \mathcal{L}(X) &= \sum_{ij} (\lambda_i+\gamma_j)\,\big\langle w_iu_j^*, X\big\rangle\; v_iz_j^* \;=\; \sum_{ij} (\lambda_i+\gamma_j)\;\big(w_i^*Xu_j\big)\; Ve_ie_j^*Z^* \\ &= V\left(\sum_{ij} (\lambda_i+\gamma_j)\;\big(w_i^*Xu_j\big)\; e_ie_j^*\right) Z^* \;=\; V\Big(\mathbf{\Pi}\circ\big(W^*XU\big)\Big)\,Z^*, \end{aligned} \tag{24}$$

[3] A simple calculation will show that $\mathcal{L}$ is normal if A and B are normal.

where $\circ$ is the Hadamard (element-by-element) product of matrices. The last equation follows from observing that W^*XU is the matrix whose ij'th entry is $\left(w_i^* X u_j\right)$. For later reference, we also calculate $\mathcal{L}^*$

$$\begin{aligned}\mathcal{L}^*(X) &= \sum_{ij}(\lambda_i^* + \gamma_j^*)\left\langle v_i z_j^*, X\right\rangle\ w_i u_j^* \ =\ \sum_{ij}(\lambda_i^* + \gamma_j^*)\ \left(v_i^* X z_j\right)\ W e_i e_j^* U^* \\ &= W\left(\sum_{ij}(\lambda_i + \gamma_j)\ \left(v_i^* X z_j\right)\ e_i e_j^*\right) U^* \ =\ W\left(\bar{\mathbf{\Pi}} \circ \left(V^* X Z\right)\right) U^*, \end{aligned} \tag{25}$$

where $\bar{\mathbf{\Pi}}$ is the complex conjugate (without transposing) of $\mathbf{\Pi}$.

The inverse $\mathcal{L}^{-1}$ (if it exists) and the pseudo-inverse $\mathcal{L}^\dagger$ can be calculated from the spectral decomposition

$$\mathcal{L}^{-1}(X) \ =\ \sum_{ij}\frac{1}{\lambda_i + \gamma_j}\left\langle w_i u_j^*, X\right\rangle\ v_i z_j^*, \qquad \mathcal{L}^\dagger(X) \ =\ \sum_{\substack{ij \\ \lambda_i+\gamma_j\neq 0}}\frac{1}{\lambda_i + \gamma_j}\left\langle w_i u_j^*, X\right\rangle\ v_i z_j^*.$$

We can also give compact formulae if we define $\mathbf{\Pi}^{-\circ}$ and $\mathbf{\Pi}^{\dagger\circ}$ using element-by-element operations

$$\left(\mathbf{\Pi}^{-\circ}\right)_{ij} \ :=\ \frac{1}{\lambda_i + \gamma_j}, \qquad \left(\mathbf{\Pi}^{\dagger\circ}\right)_{ij} \ :=\ \begin{cases} \dfrac{1}{\lambda_i + \gamma_j}, & \lambda_i + \gamma_j \neq 0, \\ 0, & \lambda_i + \gamma_j = 0. \end{cases}$$

$\mathcal{L}^{-1}$ and $\mathcal{L}^\dagger$ can now be rewritten as

$$\begin{aligned}\mathcal{L}^{-1}(X) &= V\left(\mathbf{\Pi}^{-\circ} \circ \left(W^* X U\right)\right) Z^*, \\ \mathcal{L}^\dagger(X) &= V\left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W^* X U\right)\right) Z^*. \end{aligned} \tag{26}$$

The above can be generalized using the spectral decomposition to any function f analytic in a neighborhood of the spectrum of $\mathcal{L}$ by

$$\left(f(\mathcal{L})\right)(X) \ =\ V\left(f^\circ(\mathbf{\Pi}) \circ \left(W^* X U\right)\right) Z^*,$$

where $f^\circ(\mathbf{\Pi})$ is the element-by-element application of the function f on each entry of the matrix $\mathbf{\Pi}$. For example, given a matrix differential equation

$$\dot{X} \ =\ AX + XB \ =\ \mathcal{L}(X), \qquad X(0) = \bar{X}$$

we can write the solution formally as $\left(e^{t\mathcal{L}}\right)(\bar{X})$. The formula above gives

$$X(t) \ =\ \left(e^{t\mathcal{L}}\right)(\bar{X}) \ =\ V\left(e^{\circ(t\mathbf{\Pi})} \circ \left(W^* \bar{X} U\right)\right) Z^*,$$

where $e^{\circ(t\mathbf{\Pi})}$ is the matrix whose ij'th entry is $e^{t(\lambda_i+\gamma_j)}$.

A.2 Solvability of the Sylvester Equations (12) and (13)

We begin with the equation (12) for V_1.

- We need to characterize $\mathcal{N}(\mathcal{L})$, the null space of $\mathcal{L}$. Using (14) we note that $\mathcal{L}(X) = 0$ means

  $$\forall i,j,\ (\lambda_{0i} - \lambda_{0j})\ (w_{0i}^* X e_j) = 0 \quad \Longleftrightarrow \quad \forall i \neq j,\ (w_{0i}^* X e_j) = 0 \quad \Longleftrightarrow \quad W_0^* X \ =\ D,$$

  where D is some diagonal matrix (the diagonal entries of D are the numbers $(w_{0i}^* X e_i)$ which cannot be determined from the above condition). Finally we note that W_0 and V_0 are inverses of each other, so $W_0^* X = D$ is equivalent to $X = V_0 D$ and we conclude

  $$X \in \mathcal{N}(\mathcal{L}) \quad \Leftrightarrow \quad X \ =\ V_0 D, \quad D = \text{any diagonal matrix}$$

- Similarly for $\mathcal{L}^*(X) \;=\; A_0^*X - X\Lambda_0^*$. It's spectral decomposition is

$$\sum_{ij}(\lambda_{0i}^* - \lambda_{0j}^*)\;\left(v_{0i}^* X e_j\right)\; w_{0i}e_j^*,$$

and again $\mathcal{L}^*(X) = 0$ iff $\forall i \neq j,\;\; \left(v_i^* X e_j\right) = 0$. Repeating the above argument we conclude

$$X \in \mathcal{N}(\mathcal{L}^*) \quad \Leftrightarrow \quad X \;=\; W_0 D, \;\; D = \text{any diagonal matrix}$$

- The Sylvester equation (12) is solvable iff $(V_0\Lambda_1 - A_1V_0) \quad \in \quad \mathcal{R}(\mathcal{L}) \;=\; \mathcal{N}(\mathcal{L}^*)^\perp$. Thus to verify solvability, we need to show that $V_0\Lambda_1 - A_1V_0$ is orthogonal to $\mathcal{N}(\mathcal{L}^*)$. Indeed

$$\begin{aligned}
X \in \mathcal{N}(\mathcal{L}^*) \Leftrightarrow\; & X = W_0 D, \\
\langle X, V_0\Lambda_1 - A_1V_0\rangle = \langle W_0 D, V_0\Lambda_1 - A_1V_0\rangle \;&=\; \mathsf{tr}\big(D^* W_0^* V_0\Lambda_1\big) - \mathsf{tr}\big(D^* W_0^* A_1 V_0\big) \\
&=\; \mathsf{tr}\big(D^*\Lambda_1\big) - \mathsf{tr}\big(D^*\; W_0^* A_1 V_0\big) \;=\; \mathsf{tr}\Big(D^*\; \mathsf{dg}(W_0^* A_1 V_0)\Big) - \mathsf{tr}\big(D^*\; W_0^* A_1 V_0\big) \\
&=\; \mathsf{tr}\big(D^*\; W_0^* A_1 V_0\big) - \mathsf{tr}\big(D^*\; W_0^* A_1 V_0\big) \;=\; 0,
\end{aligned}$$

where we used $\Lambda_1 = \mathsf{dg}(W_0^* A_1 V_0)$, and that D is a diagonal matrix.

For equation (13), note that by transposing it we get a Sylvester equation of a similar form to (12)

$$A_0^* W_1 - W_1\Lambda_0^* \;=\; W_0\Lambda_1^* - A_1^* W_0.$$

This equation is $\mathcal{L}^*(W_1) = W_0\Lambda_1^* - A_1^*W_0$ where $\mathcal{L}^*$ is the adjoint of the Sylvester operator of equation (12), which we have already analyzed. In particular, the spectral decomposition of $\mathcal{L}^*$ can be calculated from (25)

$$\mathcal{L}^*(Y) \;=\; W_0\left(\bar{\mathbf{\Pi}} \circ \left(V_0^* Y\right)\right).$$

Applying the pseudo-inverse to the right hand side

$$\begin{aligned}
W_1 = \;& \mathcal{L}^{*\dagger}(W_0\Lambda_1^* - A_1^*W_0) \;=\; W_0\left(\bar{\mathbf{\Pi}}^{\dagger\circ} \circ \left(V_0^*\left(W_0\Lambda_1^* - A_1^*W_0\right)\right)\right) \;=\; W_0\left(\bar{\mathbf{\Pi}}^{\dagger\circ} \circ \left(\Lambda_1^* - V_0^*A_1^*W_0\right)\right) \\
=\; & -W_0\left(\bar{\mathbf{\Pi}}^{\dagger\circ} \circ \left(V_0^*A_1^*W_0\right)\right),
\end{aligned}$$

where the last equality follows from $\bar{\mathbf{\Pi}}^{\dagger\circ}$ having all zeros on the diagonal. We can rewrite the solution for W_1^* by noting that $\bar{\mathbf{\Pi}}^* = \bar{\mathbf{\Pi}}^T = -\mathbf{\Pi}$ and similarly $\left(\bar{\mathbf{\Pi}}^{\dagger\circ}\right)^* = -\mathbf{\Pi}^{\dagger\circ}$

$$\begin{aligned}
W_1^* = \;& -\left(W_0\left(\bar{\mathbf{\Pi}}^{\dagger\circ} \circ \left(V_0^*A_1^*W_0\right)\right)\right)^* \;=\; -\left(\left(\bar{\mathbf{\Pi}}^{\dagger\circ}\right)^* \circ \left(V_0^*A_1^*W_0\right)^*\right) W_0^* \\
=\; & \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^*A_1V_0\right)\right) W_0^*
\end{aligned}$$

A.3 Background: Spectral Decomposition of a Matrix

Just as matrix partition notation is useful in expressing eigenvalue/eigenvector relations (2), it is also useful to understand diagonalization and spectral decomposition of a matrix. First observe that (2) also gives the left eigenvectors of A as follows

$$AV \;=\; V\Lambda \quad \Leftrightarrow \quad V^{-1}A \;=\; \Lambda V^{-1} \quad \Leftrightarrow \quad A^*V^{-*} \;=\; V^{-*}\Lambda^*,$$

thus rows of V^{-1} (columns of V^{-*}) are left eigenvectors of A (right eigenvectors of A^*). These equations also give the diagonalization

$$A \;=\; V\Lambda V^{-1}.$$

which can also be interpretated as a rank-1 decomposition of A as follows. Let $W^* := V^{-1}$ (or equivalently $W := V^{-*}$), and observe that the diagonalization can be rewritten as

$$A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} w_1^* \\ \vdots \\ w_n^* \end{bmatrix}$$

$$= \lambda_1 \begin{bmatrix} v_1 \end{bmatrix} \begin{bmatrix} w_1^* \end{bmatrix} + \cdots + \lambda_n \begin{bmatrix} v_n \end{bmatrix} \begin{bmatrix} w_n^* \end{bmatrix} = \sum_{i=1}^{n} \lambda_i \, v_i w_i^*. \tag{27}$$

This is a rank-1 (aka dyadic) decomposition of A, namely into n rank-1 matrices made up of outer products of the respective columns of V and W scaled by the respective eigenvalue. Note that since $W^*V = I$, there are simple relationships between the columns of V and W.

- *The sets $\{v_i\}$ and $\{w_i\}$ form a reciprocal basis:* Reciprocal bases[4] have the property that $v_i^* w_j = \delta_{i-j}$. This is easily seen to be true from the following partitioning of $W^*V = I$

$$V^{-1}V = W^*V = \begin{bmatrix} w_1^* \\ \vdots \\ w_n^* \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} w_1^* v_1 & \cdots & w_1^* v_n \\ \vdots & & \vdots \\ w_n^* v_1 & \cdots & w_n^* v_n \end{bmatrix} = I.$$

  The reciprocal basis is useful since it allows for writing any vector x in terms of a basis $\{v_i\}$ by observing

$$x = VW^*x = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix} \begin{bmatrix} w_1^* \\ \vdots \\ w_n^* \end{bmatrix} \begin{bmatrix} x \end{bmatrix} = \sum_{i=1}^{n} v_i \langle w_i, x \rangle . \tag{28}$$

  Thus the coefficients of expansion of a vector x in a basis $\{v_i\}$ are the inner products $\langle w_i, x \rangle$ of the vector with the respective elements of the reciprocal basis $\{w_i\}$.

- We can obtain yet another interpretation of the action of a diagonalizable matrix on a vector by acting with (27) on any vector x

$$Ax = \sum_{i=1}^{n} \lambda_i \, v_i w_i^* x = \sum_{i=1}^{n} \lambda_i \, v_i \langle w_i, x \rangle = \sum_{i=1}^{n} \lambda_i \, P_i x \tag{29}$$

  Each matrix $P_i := v_i w_i^*$ is a (not necessarily orthogonal) projection operator (note: $P^2 = P$) onto an individual eignsubspace of A. Note that the inner product is with the corresponding *left* eigenvector w_i. This expression above is known as the spectral decomposition of a linear operator.

- Let f be any function analytic in a neighborhood of the set $\{\lambda_1, \ldots, \lambda_n\}$, then

$$f(A) = \sum_{i=1}^{n} f(\lambda_i) \, P_i$$

Note that when A is Hermitian, columns of V are orthonormal, $W = V$, and the projections above are orthogonal.

A.4 Repeated Eigenvalues Case for Non-normal A_0

Now assume A_0 has a repeated eigenvalue $\bar{\lambda}$ (multiplicity m) and define a basis so that A_0 is block-diagonal with the first m basis elements a basis for the eigensubspace of $\bar{\lambda}$. With this basis, A_0, V_0, W_0 and Λ_0 have a block partitioning as

$$A_0 = \begin{bmatrix} (A_0)_{11} & 0 \\ 0 & (A_0)_{22} \end{bmatrix}, \; V_0 = \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix}, \; W_0 = \begin{bmatrix} (W_0)_{11} & 0 \\ 0 & (W_0)_{22} \end{bmatrix}, \; \Lambda_0 = \begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix}$$

[4] Another common term is *dual bases*.

where I is the $m \times m$ identity matrix. Now take eq. (7)

$$\Lambda_0 V_1 + A_1 V_0 \;=\; V_0 \Lambda_1 + V_1 \Lambda_0.$$

If we partition all matrices conformably with the above partitions, we get

$$\begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix} \begin{bmatrix} (V_1)_{11} & (V_1)_{12} \\ (V_1)_{21} & (V_1)_{22} \end{bmatrix} + \begin{bmatrix} (A_1)_{11} & (A_1)_{12} \\ (A_1)_{21} & (A_1)_{22} \end{bmatrix} \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix}$$
$$= \begin{bmatrix} (V_0)_{11} & 0 \\ 0 & (V_0)_{22} \end{bmatrix} \begin{bmatrix} (\Lambda_1)_{11} & 0 \\ 0 & (\Lambda_1)_{22} \end{bmatrix} + \begin{bmatrix} (V_1)_{11} & (V_1)_{12} \\ (V_1)_{21} & (V_1)_{22} \end{bmatrix} \begin{bmatrix} \bar{\lambda} I & 0 \\ 0 & (\Lambda_0)_{22} \end{bmatrix}.$$

Taking the 11 block equation we get

$$\bar{\lambda} I_n \; (V_1)_{11} \;+\; (A_1)_{11}(V_0)_{11} \;=\; (V_0)_{11}(\Lambda_1)_{11} \;+\; (V_1)_{11} \; \bar{\lambda} I_n,$$

but $\bar{\lambda} I_n$ commutes with any matrix, so we simply get

$$(A_1)_{11}(V_0)_{11} \;=\; (V_0)_{11}(\Lambda_1)_{11},$$

from which we can obtain $(\Lambda_1)_{11}$ by left multiplication by $(W_0^*)_{11}$

$$(\Lambda_1)_{11} \;=\; (W_0)^*_{11}(A_1)_{11}(V_0)_{11} \;=\; (W_0^* A_1 V_0)_{11},$$

where the last equality follows from the block-diagonal structure of V_0 and W_0. Now let's examine the meaning of this. The perturbation expansion (3) assumes that all the Λ_i's are diagonal matrices (otherwise (3) is not an eigenvalue/eigenvector relation). This equation states that for $(\Lambda_1)_{11}$ to be diagonal as required, we must choose $(V_0)_{11}$ so that its columns are the eigenvectors of $(A_1)_{11}$. Consequently, $(\Lambda_1)_{11}$ will simply be the diagonal matrix of eigenvalues of $(A_1)_{11}$.

The meaning of the above is that only this particular choice of V_0 will yield the expansion (3). Unlike the non-repeated case, we also have to discover V_0, not just V_1 and Λ_1.

Now to calculate V_1 we can apply the psuedo-inverse formula (26) (which produced (15) earlier), but now we pay special attention to the structure of $\mathbf{\Pi}^{\dagger\circ}$. In this case, it is

$$\left(\mathbf{\Pi}^{\dagger\circ}\right)_{ij} \;:=\; \begin{cases} \dfrac{1}{\lambda_{0i} - \lambda_{0j}}, & \lambda_{0i} \neq \lambda_{0j}, \\ 0, & \lambda_{0i} = \lambda_{0j}. \end{cases}$$

It is instructive to look at the structure of $\mathbf{\Pi}^{\dagger\circ}$, it is of the following form

$$\mathbf{\Pi}^{\dagger\circ} \;=\; \begin{bmatrix} 0 & & \\ & 0 & \\ & & \ddots \end{bmatrix},$$

where the large 0 block is $m \times m$, the diagonal is all zeros, and the remaining terms are $1/(\lambda_{0i} - \lambda_{0j})$. The expression for V_1 is now the same as (15), but now we also make sure to use the V_0 that made Λ_1 diagonal above (recall that the derivation of (15) assumed Λ_1 to be diagonal)

$$V_1 \;=\; -\, V_0 \left(\mathbf{\Pi}^{\dagger\circ} \circ \left(W_0^* A_1 V_0 \right) \right).$$

Again, to compare with existing expressions, find v_{1k} using $V_1 e_k$

$$v_{1k} \;=\; -\, V_1 e_k \;=\; -\sum_{\lambda_{0i} \neq \lambda_{0k}} \frac{w_{0i}^* A_1 v_{0j}}{\lambda_{0i} - \lambda_{0j}} \; v_{0i} e_j^* \; e_k \;=\; \sum_{\lambda_{0i} \neq \lambda_{0k}} \frac{w_{0i}^* A_1 v_{0k}}{\lambda_{0k} - \lambda_{0i}} \; v_{0i}, \tag{30}$$

References

[1] H. Baumgärtel, "Analytic perturbation theory for matrices and operators," *Operator theory*, vol. 15, 1985.

[2] J. J. Sakurai and E. D. Commins, "Modern quantum mechanics, revised edition," 1995.